ANÁLISIS DE LOS SISTEMAS BUSINESS INTELLIGENCE Y SU APLICACIÓN PRÁCTICA EN LOS PROYECTOS SOFTWARE

PROYECTO FIN DE CARRERA

Universidad Carlos III de Madrid

Ingeniería Técnica en Informática de Gestión

Autor	Carmen Cámara Núñez
Director	Fuensanta Medina Domínguez
Departamento	Informática
Presentación	Noviembre de 2010

Agradecimientos

A Fuensanta, por proponerme el estudio de Business Intelligence, por ayudarme en la realización del proyecto y por estar siempre disponible. Sin ella no hubiera sido posible. Mi camino en la universidad continua, y me gustaría poder volver a contar contigo. Gracias Fuensanta.

A mis padres y tía madrina, que me han apoyado en todos mis propósitos desde el principio, y sin los cuales no hubiera podido lograrlo.

A Mamel por sus largas charlas, sobre todo en momentos de indecisión.

A Toni y Pilar, por recorrer la carrera conmigo.

A Sol.

A todos, infinitas gracias.

Contenido

Índice de ilustraciones

Índice de tablas

1. Introducción

Con el aumento de las tecnologías y su aplicación al mundo empresarial, aumentan los volúmenes de datos. En un principio se consultaba a los sistemas transaccionales para recuperar la información, esta solución puede funcionar para aspectos concretos, pero estos sistemas no cubren por completo las necesidades de explotación de la información, ya que están diseñados meramente para soportar transacciones.

La información es un activo principal en las organizaciones, gestionarla de manera correcta determina el éxito o fracaso de la misma. Para ello, se hacen necesarias soluciones que provean un mejor acceso a la información y, a lo que aporta valor, a su análisis. Business Intelligence (BI) tiene el objetivo de apoyar de forma sostenible y continuada a las organizaciones facilitándolas información necesaria para apoyar sus decisiones y que de esta manera puedan ser más competitivas.

BI comprende una gran variedad de herramientas analíticas pensadas para asistir en la toma de decisiones a través de la búsqueda de conocimiento en la información de la propia organización.

Una breve evolución que lleva hasta la concepción de BI como se conoce hoy sería la que sigue:

- 1960s: Los datos se almacenaban en sistemas de archivo y se consultaban con operaciones de acceso.

- 1969: Creación del concepto de base de datos de la mano de Codd. A partir de este punto, los usuarios no necesitaban conocer los detalles del almacenamiento físico de los datos.

- 1970's: Desarrollo de las primeras bases de datos y primeras aplicaciones empresariales (SAP, JD Edwards, Siebel, PeopleSoft). Estas aplicaciones permitieron realizar "data entry" en los sistemas, aumentando la información disponible, pero aún no realizaban un acceso rápido y fácil a dicha información.

- 1980s: Creación del concepto Datawarehouse por Ralph Kimball y Bill Inmon, y aparición de los primeros sistemas de reporting. A pesar de todo, seguía siendo complicado y con una funcionalidad no muy extensa. Existían sistemas de bases de datos pero no había aplicaciones que facilitasen su explotación.

- 1989: Introducción del término Business Intelligence por Howard Dresner.

- 1990s: Nacen múltiples aplicaciones BI. Aún resultaban muy caras, pero facilitaban acceso a la información. Aparición de las bases de datos multidimensionales (cubos) y del concepto de Datamining.

- 2000s: Consolidación de las aplicaciones BI en unas pocas plataformas Business Intelligence (Oracle, SAP, IBM, Microsoft). Aparición de Knowledge Access, tecnología aún reciente y poco evolucionada.

Cada concepto que fue apareciendo, venía precedido del anterior, y se trataba de una mejora en cuanto a las carencias de éste. BI es el resultado de esta evolución, hoy en día es en las grandes empresas donde más se explotan las herramientas BI.

A medida que crece el volumen de datos y las bases de datos se hacen más extensas, la explotación de esas bases de datos se vuelve más compleja. Mediante técnicas de Datamining se puede analizar y buscar relaciones, incluso aunque éstas no sean obvias, entre variables de la base de datos. De no contar con este tipo de herramientas, esta tarea resultaría imposible. A partir de unos objetivos predeterminados, pueden buscarse patrones o reglas del dominio que trate la información.

Muchas de las técnicas que han ido surgiendo tuvieron su origen en la investigación, y comparten base teórica con algunos aspectos de la Inteligencia Artificial. Desde la perspectiva de la Inteligencia Artificial, Datamining es una etapa dentro del proceso de extracción de conocimiento en bases de datos, KDD (de sus siglas en inglés, Knowledge Discovery in Database).

Las últimas tendencias del mercado BI apuntan hacía su implementación en Internet. Ya existen herramientas que permiten un acceso vía web para promover el acceso desde cualquier lugar, facilitar el uso y minimizar los costes de administración y mantenimiento. Los grandes consumidores de estas plataformas son las grandes empresas, que suelen ser de carácter multinacional. También se empieza a aplicar BI en el comercio electrónico, para ayudar a la gestión de negocios y decisiones en Internet.

Puede creerse que BI son técnicas estadísticas, pero es mucho más. Y se demuestra cuando la información a analizar es muy dinámica, BI supera a estas aplicaciones, y proporciona información difícil de conseguir con un tiempo de respuesta muy corto.

1.1 Motivación

Hoy día gracias a internet y al desarrollo de los sistemas de información en las empresas, éstas disponen de grandes cantidades de información, el potencial que esto supone para la toma de decisiones es enorme. Pero para hacer a la empresa competitiva, a menudo se dispone de poco tiempo para tomar la decisión. Algunas decisiones tienen una importancia relativa en el desarrollo de la actividad de la empresa, pero otras son de vital importancia, y lo que interesa en cualquier caso es la toma de una decisión correcta.

Para los jefes de proyecto o responsables al efecto de generar información directiva, directivos en general, el proceso de toma de una decisión es una de las mayores responsabilidades. En el mundo empresarial se suele decir que las decisiones son algo así como el motor de los negocios, y en efecto, de la adecuada selección de alternativas depende en gran parte el éxito de la organización.

Pero en un mundo cada vez más complejo y en continuo cambio, la necesidad de tomar decisiones rápidamente puede llegar a ser desconcertante. En este escenario los directivos se encuentran desbordados con peticiones de información urgente,

continua y no coordinada, sobre la que tienen que decidir. Algo peor que no tener información es no saber qué hacer con ella.

Se hace necesaria la ayuda de sistemas que den soporte a la toma de decisiones, presentando la información de una manera sencilla e intuitiva y permitan al usuario, de un solo vistazo, poseer la información que necesita en cada momento.

Un artículo [Rice, 2004] sobre el sector sanitario en EE.UU. decía lo siguiente:

> *"En muchos hospitales los analistas financieros destinan un 80% del tiempo a agregar y normalizar manualmente información en hojas de cálculo Excel, y tan sólo un 20% a analizar la información relevante".*

Afirmaciones como ésta deberían dar a las organizaciones los ánimos suficientes para comenzar el camino hacia la implementación de proyectos de *Business Intelligence*. No es admisible que se invierta el 80% del tiempo preparando información y tan sólo el 20% analizándola: los porcentajes deberían ser, obviamente, al revés, es decir, destinar la mayor parte del tiempo a analizar la información y tan sólo una pequeña parte del tiempo a prepararla. Lo que realmente tiene valor en la organización es la toma de decisiones, no la preparación de la información.

Esta situación resulta razón suficiente y de tremendo interés para explorar el mundo del BI y encontrar de qué manera puede facilitarnos la toma de decisiones para manejar nuestros proyectos o negocios.

Según *Gartner Research* [Gartner], la falta de conocimiento es la mayor amenaza de las organizaciones modernas y señala *"el objetivo del Business Intelligence es eliminar las conjeturas y la ignorancia en los ambientes empresariales, aprovechando los vastos volúmenes de datos cuantitativos que las empresas recolectan todos los días en sus diversas aplicaciones corporativas".*

Por otro lado, hoy en día, el abaratamiento de los datawarehouse y la aparición de nuevas herramientas BI, que han abierto la veda para una competición más dura en este sector, incluso la aparición de herramientas opensource, han hecho que los sistemas BI se encuentren al alcance de más organizaciones, convirtiéndose en herramientas populares de gran utilidad y uso.

Con este proyecto se pretende conocer el mundo del Business Intelligence, las tecnologías que lo soportan y ver qué valor puede aportar realmente en la mejor toma de decisiones.

1.2 Descripción general del proyecto

En el presente proyecto se expondrá una visión amplia y actual sobre Business Intelligence como principal tendencia como sistema de apoyo a las decisiones. Se realizará también una aplicación BI utilizando una de las plataformas más relevantes del mercado, *QlikView.*

No es objetivo de este proyecto analizar los algoritmos usados en BI, encargados de explorar en los datos y cuyo resultado es el conocimiento que se busca. La idea es realizar un estudio sobre Business Intelligence, qué es, dónde se aplica, qué se necesita para su implementación, cómo se desarrollan estos proyectos y qué tipo de herramientas hay en el mercado actual.

A lo largo del presente documento se procurará la consecución de lo siguiente:

- Definir el concepto de Business Intelligence y justificar su necesidad por parte de las organizaciones.
- Definir la arquitectura de los sistemas Business Intelligence, estudiando los diferentes elementos que lo componen.
- Conocer la gestión de proyectos Business Intelligence.
- Conocer el uso de BI y futuras tendencias.
- Realización de una aplicación BI.

La estructura del documento está dividida en tres capítulos principales

- Capítulo 1. Introducción a BI, conceptos generales.
- Capítulo 2. Estudio de Business Intelligence, que contendrá aspectos más específicos como: Definición de BI, motivos para su uso, arquitectura de las aplicaciones BI, herramientas, gestión de proyectos BI, y uso y tendencias.

- Capítulo 3. Aplicación BI: Justificación de la elección de la herramienta escogida para el desarrollo de la aplicación BI realizada, arquitectura de dicha herramienta y explicación de la aplicación en un manual de usuario.

- Capitulo 4. Planificación y recursos: Donde se expondrán los recursos utilizados y la planificación seguida en el proyecto.

- Capitulo 5. Conclusiones observadas de la realización del proyecto y propuesta de futuros trabajos.

- Anexo DIKW. Donde se explicará la diferencia entre datos, información, conocimiento y sabiduría.

1.3 Ámbito

Podríamos decir que Business Intelligence pertenece a dos ámbitos profesionalmente diferenciados:

- Por un lado pertenece al ámbito de la empresa puesto que consiste en una plataforma que aporta apoyo a las decisiones de negocio del ámbito empresarial. En este sentido los usuarios beneficiarios, son por lo tanto las organizaciones, es a ellas a las que van destinadas las plataformas BI y es en donde mayor grado de explotación pueden obtener éstas.

  Se tratará el BI desde esta cara cuando se explica cómo gestionar los proyectos BI, cómo seleccionar la herramienta adecuada, qué son los modelos y los indicadores del negocio, etc.

- Y por otro lado debemos considerarlo también del ámbito de la Ingeniería informática, ya que se tratan conceptos de datamining, datawarehouse, datamart, etc. Y es desde la informática donde sale la implementación de estas plataformas.

  Se detallará a lo largo del presente proyecto cómo se implanta una solución BI, cómo se gestionan los proyectos BI, qué herramientas BI hay ahora mismo en el mercado y cómo pueden éstas clasificarse, etc.

2. Estudio de Business Intelligence

2.1 Definición de Business Intelligence

En este apartado se expondrá una definición de BI más formal y avanzada, a la dada de manera general en el capítulo anterior.

Fue Howard Dresner, posterior consultor de *Gartner Research*, el primero que acuñó en 1989 el término de Business Intelligence (BI). Éste, definió BI como:

> 'BI es un proceso interactivo para explorar y analizar información estructurada sobre un área (normalmente almacenada en un datawarehouse), para descubrir tendencias o patrones, a partir de los cuales derivar ideas y extraer conclusiones. El proceso de Business Intelligence incluye la comunicación de los descubrimientos y efectuar los cambios. Las áreas incluyen clientes, proveedores, productos, servicios y competidores."

Desglosando esta definición:

- *Proceso interactivo* porque el BI es un proceso de análisis de información continuo, no sólo en un momento puntual. Business Intelligence permite contrastar la situación actual con una situación retrospectiva de años atrás.

- *Explorar,* ya que en el momento inicial en que nos encontramos por primera vez con la información que BI nos presenta, "exploramos" los datos de los que disponemos e incluso es posible que descubramos relaciones que no conocíamos.

- *Analizar*, ya que el usuario lo que pretende es descubrir relaciones entre variables, así si descubrimos una fuerte relación entre tipo y modelo y patrones utilizados, esto nos dará conocimiento acerca de nuestra elección de patrones para el proyecto que nos ocupa

- *Información estructurada y datawarehouse,* los datos de los que se sirven las plataformas BI se encuentran almacenadas en tablas relacionadas entre sí, y estas tablas se encuentran almacenadas en datawarehouses.

- *Área de análisis,* porque todo proyecto BI debe estar centrado en un área en concreto.

En conclusión y con palabras más coloquiales llamamos BI al conjunto de estrategias y herramientas que permiten reunir, depurar y transformar datos de los sistemas transaccionales[1], y normalmente desestructurados, en información estructurada para su explotación directa (reporting, análisis OLTP/OLAP…) o para analizarla y así convertirla en conocimiento, teniendo así el soporte a la toma de decisiones del negocio.

Es un conjunto de estrategias y herramientas dirigidas a convertir datos en información e información en conocimiento mediante el análisis del conjunto de datos que se disponen en una organización, independientemente de cuál sea su soporte, para conducir de manera eficaz los procesos de negocio de las organizaciones.

En la actualidad Business Intelligence constituye una disciplina puntera para obtener conocimiento de un negocio.

[1] Un sistema transaccional es un tipo de sistema de información diseñado para recolectar, almacenar, modificar y recuperar todo tipo de información que es generada por las transacciones en una organización.

Los distintos usos que soportan las herramientas BI son:

- Reporting: Elaboración de informes, el reporting es considerado el núcleo de BI, y es la aplicación de BI más utilizada.
- Análisis: Herramientas de consulta ad-hoc y OLAP.
- Planificación y modelización.
- Monitoring: Dashborads y Scorecards.
- Análisis avanzado: Datamining y Textmining.

Los primeros sistemas BI eran herramientas individuales especializadas en uno de estos usos o funciones. Hoy día las soluciones BI son herramientas completas que realizan varias de estas funcionalidades.

Incluso los sistemas BI modernos no son sólo soluciones que navegan por los datos, son también herramientas de desarrollo y administración, que permiten crear, desarrollar y mantener aplicaciones BI.

A continuación se detallarán en mayor profundidad algunos de los usos más comunes de las herramientas BI.

2.1.1 Cuadros de Mando Integrales (CMI)

El cuadro de mando integral, en inglés Balanced Scorecard (BSC), fue presentado por primera vez por Rober Kaplan y David Norton en la revista Havard Business Review en 1992.

El CMI es una herramienta para medir las actividades de una organización en términos de su planificación estratégica, que muestra continuamente cuándo la compañía alcanza los resultados definidos en su plan estratégico. También es una herramienta que ayuda a expresar los objetivos e iniciativas necesarias para cumplir con la estrategia.

El CMI está orientado al seguimiento de unos indicadores, los que la empresa defina. Estos indicadores se extraen del plan estratégico de la empresa, y se muestran de manera gráfica y continua en el CMI, proporcionando información en cada momento

del punto en el que se encuentra la empresa, conforme a los objetivos o metas planificados en relación a dichos indicadores.

Por otro lado, un CMI es controlado por la dirección general de una compañía, frente a otras funcionalidades de Business Intelligence más enfocadas a la dirección departamental.

En el libro 'The Balanced ScoreCard: Translating Strategy into Action' se encuentra la siguiente definición, bastante precisa, de los cuadros de mando integrales:

> "el BSC es una herramienta revolucionaria para movilizar a la gente hacia el pleno **cumplimiento de la misión**, a través de canalizar las energías, habilidades y **conocimientos** específicos de la gente en la organización hacia el logro de metas estratégicas de largo plazo. Permite tanto guiar el desempeño actual como apuntar el desempeño futuro. Usa medidas en cuatro categorías -desempeño financiero, conocimiento del cliente, procesos internos de negocios, y aprendizaje y crecimiento- para alinear iniciativas individuales, organizacionales y tras-departamentales e identifica procesos enteramente nuevos para cumplir con objetivos del cliente y accionistas. El BSC es un robusto sistema de aprendizaje para probar, obtener realimentación y actualizar la **estrategia de la organización.** Provee el sistema gerencial para que las compañías inviertan en el largo plazo en clientes, empleados, desarrollo de nuevos productos y sistemas más bien que en gerenciar la última línea para bombear utilidades de corto plazo. Cambia la manera en que se mide y gerencia un negocio".

Las principales ventajas que presenta el uso de un CMI son:

- Poder traducir el plan estratégico de la organización a una serie de indicadores facilitando el consenso sobre los propios objetivos de la empresa y sobre cómo alcanzarlos.

- Deja ver cómo las diferentes acciones que se llevan a cabo en la empresa, tanto a corto como a largo plazo, influyen en los indicadores, y por lo tanto en los objetivos de la organización.

- Permite identificar desviaciones en el plan estratégico y descubrir qué originó estas desviaciones.

- Puede utilizarse además, como panel para comunicar el plan estratégico de la empresa.

No obstante la implantación de un CMI conlleva inconvenientes, éstos son:

- Realizar un modelo de cuadro de mando poco elaborado o que no cuente con la colaboración de la dirección, no aportará valor alguno.

- Han de saber escogerse los indicadores, porque si no se escogen los adecuados, el CMI pierde una buena parte del valor que debe aportar, porque no representará fielmente el plan de la empresa y por lo tanto no servirá al fin que tiene la herramienta.

- Debe estar actualizado, si no se adapta conforme cambian los objetivos de la empresa, se encontrará desfasado y será inútil.

- Es contraproducente utilizar un CMI como herramienta de control de la estrategia de la empresa se encuentra todavía en evolución. En este caso podría ser utilizado, más bien, como herramienta de aprendizaje.

2.1.2 Sistemas de Soporte a la Decisión (DSS)

El principal objetivo de los Sistemas de Soporte a la Decisión (del inglés, *Decision support system*) es explotar al máximo la información residente en la organización de la empresa, sea cual sea su soporte. Aquí encontramos la diferencia fundamental con los CMI que se centran en representar básicamente indicadores de la organización. Otra diferencia, ya comentada, es que los DSS, por su utilidad, van dirigidos a la gerencia departamental y no a la alta dirección de la empresa.

En términos bastante más específicos, un DSS es "un sistema de información basado en un computador interactivo, flexible y adaptable, especialmente desarrollado para apoyar la solución de un problema de gestión no estructurado para mejorar la toma de decisiones. Utiliza datos, proporciona una interfaz amigable y permite la toma de decisiones en el propio análisis de la situación" [Turban, 1995]

Se puede pensar que el análisis de datos es un proceso sencillo, y fácil de llevar a cabo mediante una aplicación hecha a medida o un ERP sofisticado. Pero no es así, estas soluciones suelen disponer de una serie de informes predefinidos presentando la información de manera estática, no permitiendo profundizar en los datos, navegar por ellos, manejarlos desde distintas perspectivas, etc.

Los sistemas de soporte a la decisión es la herramienta BI más utilizada por sus características:

- Informes dinámicos, flexibles e interactivos: El usuario decide en cada momento que información quiere que la herramienta le brinde, y conforme a esto la utiliza, de manera que no tiene que ceñirse a los listados predefinidos que se configuraron en el momento de la implantación, y que no siempre responden a sus dudas reales.

- No requiere conocimientos técnicos: Para examinar la información disponible o crear nuevas consultas no es imprescindible tener conocimientos técnicos al efecto o buscar ayuda en el departamento de informática.

- Rapidez en el tiempo de respuesta, ya que la herramienta BI cuenta con una base de datos por debajo que suele ser un datawarehouse corporativo y/o un datamart, ambas estructuras optimizadas para el análisis de grandes volúmenes de información (Se ampliará este tema en el apartado de Arquitectura del presente documento)

- Integridad referencial absoluta: Se denomina así a la Integración entre todos los sistemas/departamentos de la compañía. El proceso de ETL previo a la implantación de un Sistema de Soporte a la Decisión (y que se explica en el apartado de arquitectura del presente documento) garantiza la integración de los datos entre las diferentes unidades de la empresa. De manera que mediante un sistema DSS podemos reunir toda la información de la compañía en un solo almacén, independientemente del formato en que ésta se encuentre.

- Cada usuario puede disponer de información adecuada a su perfil: Puede usarse la herramienta de manera que cada rol tenga acceso a la información que necesita para mejorar la eficiencia en su trabajo, y no a más.

- Disponibilidad de información histórica: Suele ser muy interesante para las empresas comparar información extraída de datos actuales con información histórica de la compañía, con el fin de analizar tendencias, fijar parámetros de negocio, etc. Las herramientas BI almacenan dicha información para que esto sea posible.

En la tabla 1 se exponen las principales diferencias entre un CMI y un DSS.

	CMI	DSS
Objetivo	Seguimiento de indicadores	Análisis de la información
Usuario principal	Dirección general de la compañía.	Dirección departamental.
Arquitectura	_	Suele necesitar motor OLAP por debajo

Tabla 1. Comparación CMI-DSS

2.2 Por qué Business Intelligence

Existe la necesidad de tomar decisiones rápidas, como se vio en la introducción, para lograr que las organizaciones sean competitivas, y estas restricciones de tiempo limitan a las personas en su capacidad de tomar buenas decisiones. Sin embargo SGBD[2] e incluso herramientas ERP[3],CRM [4] o SCM [5] sofisticadas que se utilizan en las empresas, suelen tener una estructura bastante inflexible para este fin, porque aunque se adaptan, en mayor o menor medida, a permitir manejar los datos, no nos permiten extraer información de ellos y mucho menos extrapolar conocimiento (se incluye un anexo sobre la diferencia de datos, información y conocimiento).

Encontramos entonces que las herramientas tradicionales tienen limitaciones:

- **Rigidez a la hora de visualizar los datos.** El usuario tiene que ceñirse a vistas/informes predefinidos que se hicieron en tiempo de diseño. De manera que si el usuario desea obtener nuevos informes, necesitaría poseer conocimientos técnicos para realizar consultas que puedan interrogar a la base de datos o, de lo contrario, se vería obligado a acudir a los departamentos de sistemas de información de su empresa.

- **Largos tiempos de respuesta.** Normalmente consultas complejas requieren la unión de tablas y si éstas tienen gran volumen de información se producen esperas que dificultan a los usuarios la fluidez de su trabajo.

- **Lagunas de datos.** Las organizaciones suelen tener sus datos normalmente representados en diferentes sistemas de información y añadidos

[2] **SGBD:** Sistema gestor de base de datos: es un tipo de software, dedicado a servir de interfaz entre la base de datos, el usuario y las aplicaciones que la utilizan.

[3] **ERP:** del inglés *Enterprise Resources Planning*: son sistemas de información gerenciales que integran y manejan muchos de los negocios asociados con las operaciones de producción y de los aspectos de distribución de una compañía comprometida en la producción de bienes o servicios.

[4] **CRM:** del inglés *Customer Relationship Management:* Sistemas informáticos de apoyo a la gestión de las relaciones con los clientes, a la venta y al marketing.

[5] **SCM:** del inglés *Supply Chain Management:* es el proceso de planificación, puesta en ejecución y control de las operaciones de la red de suministro con el propósito de satisfacer las necesidades del cliente con tanta eficacia como sea posible

en momentos diferentes, por lo tanto no suelen tener la información integrada, lo que supone un problema a la hora de tener que utilizarla para decidir.

Por poner un ejemplo de este aspecto, en ocasiones se dan situaciones en empresas tales como que el día de una reunión en la que se tiene destinado hablar sobre el porcentaje de ventas de las diferentes sedes de la compañía, cada uno de los directivos se presenta con datos distintos, en estos términos es imposible tomar decisiones de valor para la empresa.

Vemos entonces que los sistemas de información tradicionales que poseen las empresas, no las están ayudando a tomar decisiones rápida y eficazmente.

Para superar estas deficiencias BI se apoya en herramientas que permiten la extracción, depuración y almacenamiento de toda la información de la organización, proporcionando posteriormente una velocidad de recuperación de la misma adecuada para generar conocimiento y ayudar a tomar decisiones rápidamente, lo cuál es el objetivo que compete a Business Intelligence.

En definitiva una solución BI permite **observar** qué está ocurriendo, **comprender** por qué ocurre, extraer conocimiento de estas informaciones y así poder **predecir** qué ocurrirá y **decidir** entonces qué camino debemos seguir.

Las plataformas BI son mejor solución a las herramientas tradicionales en dos aspectos:

BI como solución tecnológica:

- BI permite centralizar, depurar y afianzar los datos: ya que lo que se hace con la plataforma BI es reunificar y homogeneizar mediante un almacén de datos, pudiendo explotar éste sin esfuerzo. De esta manera todas las personas poseen la misma información, siempre actualizada, y pueden acceder a ella de una manera más rápida y sencilla.

- Ayuda a descubrir información que no es evidente para las aplicaciones tradicionales, porque la manera de presentar la información de las

herramientas BI nos ayuda a descubrir pautas, tendencias, cambios, casi imposibles de descubrir cuando no tenemos la información centralizada o se encuentra mal relacionada. Todo esto con un simple vistazo a la aplicación.

- Permite a usuarios con perfil no técnico construir nuevos informes y navegar por los datos de la compañía, con el objetivo de descubrir información que les resulte relevante.

- Y por último ayuda a optimizar el rendimiento de nuestros sistemas, porque las herramientas BI están diseñadas para perfeccionar las consultas de alto nivel, realizando las transformaciones oportunas a cada sistema (OLTP - OLAP), y liberando los servidores operacionales.

BI como ventaja competitiva:

Hay una serie de características que nos aporta BI y que al final constituyen una ventaja competitiva para nuestro proyecto/negocio.

- La primera es que permite hacer un seguimiento real de nuestros planes estratégicos. Así si disponemos de proyectos en marcha, BI nos ayudará, con la utilización de un cuadro de mando, a crear, manejar y monitorizar las métricas y objetivos presentes del proyecto.

- Además da la posibilidad de aprender de los errores cometidos en el pasado, ya que se realizan históricos de los datos relevantes. Ayudando así a no tropezar dos veces en la misma mala decisión o acción.

- Se obtiene el verdadero valor de las aplicaciones de gestión: Durante mucho tiempo las empresas han explotado con dificultad y ayudados por técnicos sus bases de datos, pero con una plataforma BI todos los empleados, desde directores hasta analistas, podrán manejar la información adecuada, integrada y actualizada pudiendo así tomar decisiones adecuadas en el momento adecuado sin que empleados subordinados tengan que estar preguntando a sus directivos constantemente por información.

BI es un sistema informático utilizado para servir de apoyo, más que automatizar, el proceso de toma de decisiones. El apoyo a la decisión significa que el sistema puede ayudar a las personas a reunir información, para que puedan generar alternativas y así tomar decisiones.

Según la consultora internacional Gartner, 7 de cada 10 compañías realizan análisis sobre sus datos de forma diaria, o incluso instantánea (Gartner, 2006). Podemos imaginar lo que las plataformas BI podrían ayudarlas cada día.

2.3 Arquitectura

En la figura 1 se ilustran los componentes que conforman un sistema BI, éstos son:

- Fuentes de información, que nutrirán de datos al sistema.
- Proceso ETL, que extrae, transforma y carga los datos de las fuentes de información al datawarehouse.
- Datawarehouse, que reúne los datos de la organización. Es una colección de datos integrada, esto es, contiene los datos de todos los sistemas operacionales de la organización, y es no volátil, la información no se modifica ni se elimina.
- Motor OLAP que permite capacidad de cálculo, consultas, análisis de escenarios en grandes volúmenes de datos
- Herramientas Front-end, que son las herramientas de visualización que permiten la navegación por los datos para su análisis por parte del usuario final.

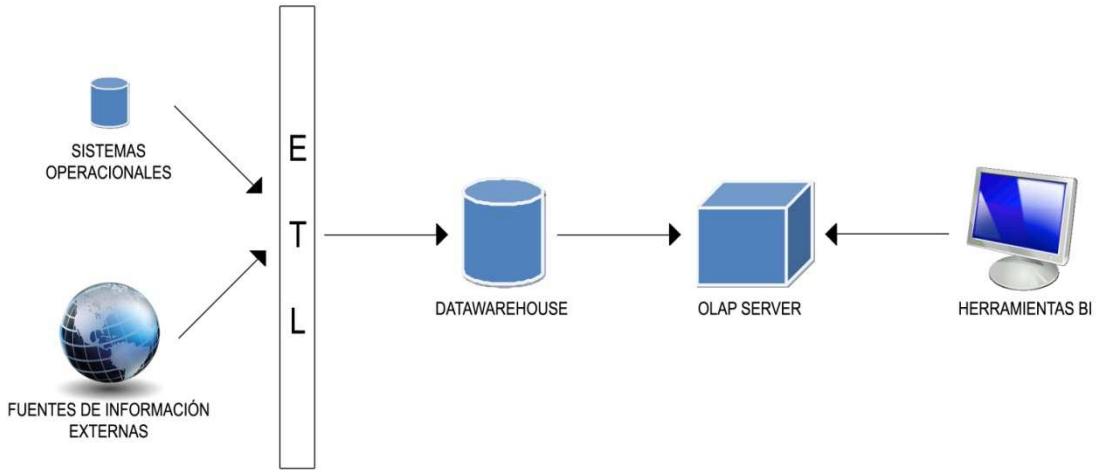


Ilustración 1. Arquitectura

A continuación se detallarán cada uno de los componentes:

2.3.1 Fuentes de información

Las fuentes de información a las que podemos acceder se clasifican en tres tipos:

- Sistemas operacionales, o transaccionales que incluyen aplicaciones desarrolladas a medida, ERP, CRM, SCM, etc.

- Sistemas de información departamentales: componen diferentes archivos más 'informales' en los que los diferentes departamentos pueden almacenar información que también puede ser interesante almacenar en el datawarehouse según el modelo de negocio: presupuestos, hojas de cálculo, ficheros planos...

- Fuentes de información externas, que constituyen un origen de datos importante para enriquecer la información de la organización, por ejemplo se podrían incorporar estadísticas del Instituto Nacional de Estadística que completarán la información de la que se dispone.

Es primordial identificar de qué fuentes vamos a recuperar la información, en qué formato se encuentra y qué calidad tiene esta información. Normalmente la información que se carga en un datawarehouse es estructurada, es decir, va a poder almacenarse en tablas, aunque también podría ser información no estructurada como informes, vídeos, correos electrónicos,etc.

Y por supuesto, es importante contrastar que la información que se dispone en las fuentes de la organización, es la que se necesita para alimentar los modelos de análisis o de negocio (El concepto de modelo de negocio se explica en el apartado 2.4 del presente documento). Si no se dispone de la información necesaria habría que llevar a cabo una modificación de las fuentes de datos transaccionales de la empresa para que éstas empezaran a recogerla.

Calidad en los datos

Es necesario verificar que la calidad de los datos es máxima, ya que si en el datawarehouse se cuelan errores, éstos se propagarán por toda la organización.

> 'Las organizaciones actúan bajo la suposición de que la información de la que se dispone es precisa y válida. Si la información no es válida,

entonces no pueden responder de las decisiones basadas en ella' [B. Inmon, 2006]

Los errores en los datos pueden desencadenar que se tomen decisiones erróneas y en consecuencia que se lleven a cabo acciones que comprometan los resultados de la organización. Los gastos que puede ocasionar el hecho de que los datos no sean de calidad pueden ser muy elevados.

'Asumir que la calidad de los datos es buena puede ser un error fatal en los proyectos BI' [M.Beyer et al, 2006]. Se debería establecer entonces unos puntos de control dentro del proyecto BI, en los que verificar la calidad de los datos, para evitar su carga en el datawarehouse.

La mayoría de los errores pueden venir de los sistemas transaccionales, por ejemplo por dejar a los usuarios introducir datos sin darles a elegir entre valores, por ejemplo. Los fallos en este punto deben llevar a una modificación de las fuentes, que a la larga resultará menos costoso que ir corrigiendo estos errores en el proceso ETL

Los errores pueden encontrarse también en el proceso ETL o en el datawarehouse.

Los principales problemas que provoca una baja calidad en los datos son [Wayne W., 2002] los que siguen y que ilustran en la figura 2.

- P1. Tiempo extra para conciliar los datos.
- P2. Baja credibilidad en el sistema.
- P3. Costes extra.
- P4. Insatisfacción de los clientes.
- P5. Retraso en el desarrollo de nuevos sistemas.
- P6. Pérdidas de ingresos.
- P7. Problemas de conformidad.

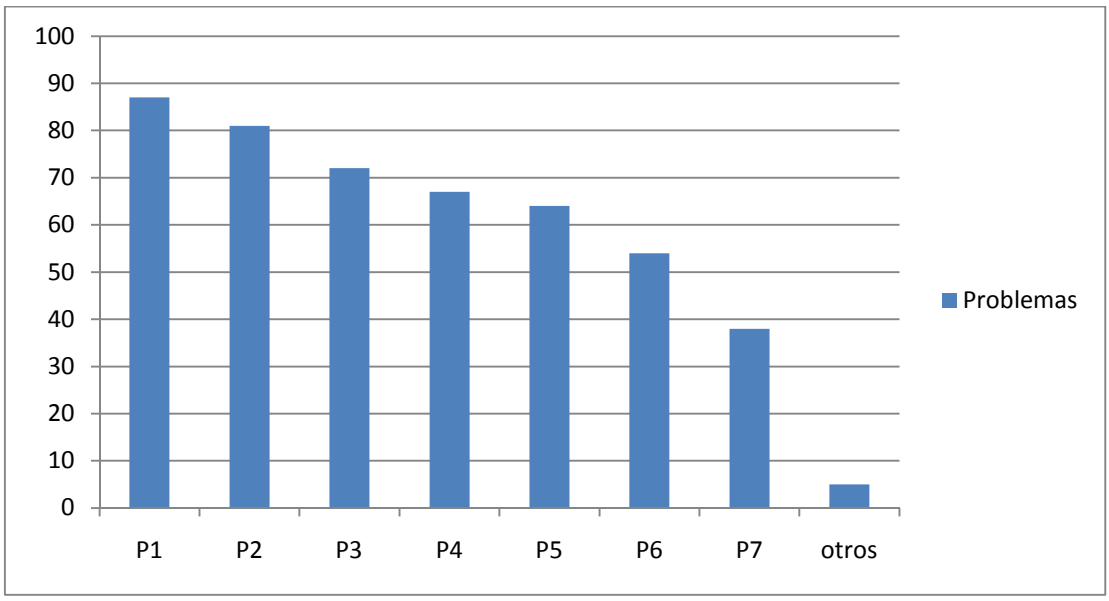


Ilustración 2. Problemas de calidad en los datos

En contra partida las ganancias que aportan la calidad en los datos son:

- Incrementos en la satisfacción de los clientes.
- Mayor confianza en los sistemas de análisis.
- Reducción de costes.
- Menor dedicación de recursos a la conciliación de datos.
- Incremento de ingresos.

Para que los datos se consideren de calidad se debe comprobar que cumplen con las siguientes características:

1. Precisión: Los datos deben representar con precisión la realidad o lo almacenado en una fuente de datos.
2. Integridad: La estructura entre los datos y sus relaciones son invariables.
3. Totalidad: Deben estar todos los datos necesarios.
4. Coherencia: Datos bien definidos y comprensibles.
5. Validez: El valor de los datos debe encontrarse entre los rangos aceptables según las normas de la organización.

6. Accesibilidad: Resulta fácil acceder a los datos.

7. Disponibilidad: Los datos se encuentran disponibles cuando se necesitan.

Para contar con una buena calidad en los datos se deben conocer los datos, saber dónde se encuentran exactamente y cuál es su nivel de calidad, para lo cual sería muy útil realizar una auditoría.

Se debería también formar un programa de calidad en el que se describieran las acciones para garantizar la calidad de los datos en la totalidad de la organización y en el que se definieran políticas y métricas para medirla.

Se podrá considerar que un datawarehouse contiene datos de calidad si no contiene datos inconsistentes ni redundantes.

Se ha de tener claro que la calidad de los datos depende tanto de los departamentos de tecnología, ya que son éstos los que conocen el significado de los datos dentro del sistema, como de la organización en sí ya que éstos conocen la semántica de los datos.

Como conclusión y resumiendo:

Cuando se seleccionen los datos que van a usarse como soporte a las decisiones debe tomarse en consideración si tienen la suficiente calidad, de lo contrario habría que determinar que acciones llevar a cabo para dársela, habría que limpiarlos (como suele decirse).

2.3.2 Proceso ETL

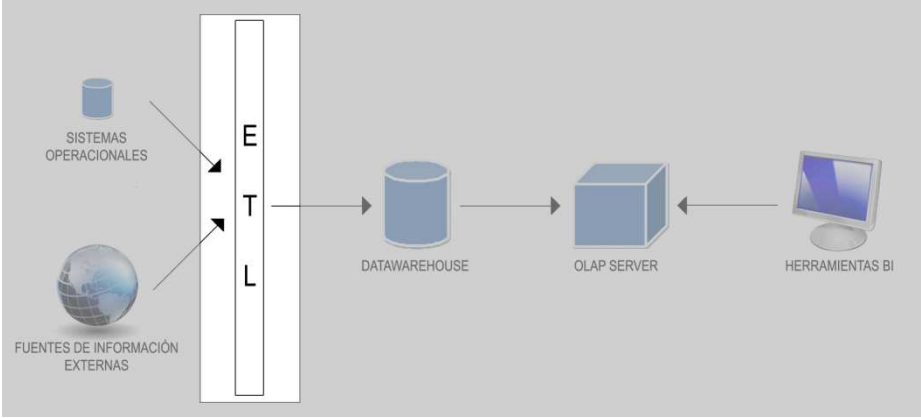


Ilustración 3. Arquitectura-Proceso ETL

La migración de datos de las fuentes de datos al datawarehouse requiere un proceso para extraer, transformar y cargar dichos datos. Este proceso se conoce como ETL, del inglés Extract, Transform and Load.

Esta parte del proceso es la más costosa, pero es necesaria por la necesidad de conciliar y limpiar los datos de origen.

El proceso ETL consta de cinco subprocesos: Extracción, limpieza, transformación, integración y actualización. Se detallan a continuación.

Extracción. Es el subproceso ETL encargado de recuperar los datos de las fuentes de información. Se extraen los datos con el uso de una herramienta ETL, pero se podrían extraer también de manera manual, con lenguajes de programación como COBOL[6].

Lo más sencillo sería duplicar los datos de las fuentes de información y pasar este conjunto a la herramienta ETL, pero de esta forma se hace trabajar a la herramienta con enormes cantidades de datos, cuando en realidad no se necesitarán todos.

La manera óptima sería limpiar, filtrar y ordenar los datos requeridos al datawarehouse, y si es posible en un sólo paso. Pero ésto cuenta con el inconveniente de que es posible que tuvieran que desmontarse las bases de datos de la organización durante unas horas, teniendo que parar las actividades que se estén realizando en ese

[6] COBOL: De las siglas ***CO**mmon **B**usiness -**O**riented **L**anguage*, lenguaje creado en 1960 con el objetivo de crear un lenguaje de programación universal que estuviera orientado principalmente a los negocios, es decir, a la llamada informática de gestión.

momento contra las fuentes de información. Esta es la razón por la que se divide el proceso en extracción, transformación y carga.

En este subproceso, extraeremos los datos necesarios y sólo los necesarios de las fuentes de información. Para ello tenemos que convenir cuáles son las mejores fuentes, de entre las que se cuentan.

Terminado de ejecutar el subproceso contamos con los datos originales.

Ha de tenerse en cuenta que si no todos los datos origen son actuales, es decir, si queremos incluir en el proceso ETL datos origen históricos, será necesario aplicar algunos cambios a los programas encargados de la carga inicial, porque al tratarse de datos antiguos es posible que no se solapen muy bien con los actuales, dando lugar a muchos posibles rechazos de datos que quieren cargarse en el proceso de limpieza.

Limpieza. Este subproceso puede considerarse un paso intermedio entre el subproceso de extracción y el de transformación.

Terminado el subproceso anterior, se poseen los datos en bruto, éstos deben ser limpiados antes de ser cargados en el datawarehouse.

Hay una serie de problemas que pueden darse cuando se extraen los datos de las fuentes originales:

- Claves primarias inconsistentes: Es posible que las claves primarias que se utilizan en las fuentes de información de origen no sean las que están definidas en el datawarehouse. Así si por ejemplo tenemos cinco ficheros de personas, cada una con su clave primaría, deben consolidarse en una única clave.

- Valores inconsistentes: Puede que existan duplicados en las fuentes, que originalmente lo eran, pero que al realizar modificaciones sobre alguna de las copias dejaron de serlo. Se deben conciliar ambos valores.

- Datos con diferentes formatos: Los campos como el campo fecha por ejemplo, admiten varios formatos posibles, se deben aplicar rutinas de estandarización para convertir todos los registros que se tengan de este campo, a un único formato, que será el que se almacene en el datawarehouse.

- Valores erróneos: Estarán presentes en las fuentes de información campos cuyos valores no serán correctos. Para su corrección la herramienta ETL aplica algoritmos de corrección en los que realiza cálculos y búsquedas en tablas e incluso en fuentes de datos externas para corregir, en la medida de lo posible, estos valores. Un ejemplo sería el algoritmo que comprobaría si un DNI entrante es o no válido.

- Datos redundantes: No siempre son fáciles de detectar, ya que el mismo nombre en varías fuentes pudiera referirse a elementos distintos, o al revés el mismo elemento de datos puede llamarse diferente en distintas fuentes.

Al finalizar este paso, se cuenta con los datos limpios.

Transformación. Después de limpiar los datos se procede al paso de transformación, que está destinado a la integración, derivación, agregación y totalización de los datos.

- Integración. Cada elemento debe conocerse con un nombre estándar. Renombraremos, de ser necesario, conforme al estándar que se siga en el datawarehouse. Así se renombrará por ejemplo DNI a DNI_Usuario.

- Derivación. Partiendo de datos individuales se calcularán nuevos datos, mediante la realización de cálculos o búsquedas. Un ejemplo sería a partir de dos datos numéricos calcular un total.

  Opuestamente, también puede crearse un nuevo dato a partir de un solo dato de origen, como por ejemplo podrían crearse los campos, día, mes, año, hora, minuto, y segundo, a partir de un campo timestamp[7]

Pero la parte primordial de la transformación se basa en el pre-cálculo de datos, para optimizar las futuras consultas al datawarehouse.

- Totalización. Se basa en el procesamiento de datos de origen de carácter numérico para obtener otro valor numérico que corresponderá a promedios, totales, máximos, mínimos, etc.

[7] Timestamp: Es una secuencia de caracteres, que denotan una hora y una fecha.

- Agregación. Consiste en crear datos a partir de tuplas de dos tablas diferentes, por ejemplo a partir de la tabla ‘usuario’ y de la tabla ‘cargos’, crear un historial puestos que ha desempeñado el usuario en la empresa.

Como es evidente, los datos se totalizan o agregan en base a la información que se quiera reportar.

Por otro lado, los datos deben conciliarse antes de cargarse en el datawarehouse. Conciliar los datos significa comprobar que la información que entra en el proceso ETL coincide con la que sale del mismo. Cada programa que lea datos y los escriba luego debe crear lo que se llaman totales de conciliación. Existen tres tipos de totales:

- Cantidad de registros. Para que las cifras coincidan, el número total de registros que entran debe ser igual al número de registros que salen más el número de registros que se rechazan, suponiendo que algún registro en el proceso de limpieza sea rechazado y no sea añadido. Este tipo de conciliación se debe llevar a cabo cuando se han extraído y mezclando filas de datos.

- Totales. Versa lo mismo que el primer tipo, pero en este caso lo que se suma y totaliza es una determinada columna de entrada. Es decir si en el tipo anterior lo que se contabilizaba era el número de campos de una tabla, aquí se contabiliza el número de atributos de la misma.

- Cantidad de dominios. La cantidad de registros para un determinado dominio (o valor) del campo de entrada, debe ser la misma que el número de registros para el mismo dominio en la salida.

Integración. Este proceso se encarga de la carga de los datos resultantes en el datawarehouse. Es un proceso sencillo, pero se deben tener en cuenta dos aspectos:

- Una vez cargados los datos en el datawarehouse debe activarse la integridad referencial para evitar que los datos almacenados se corrompan.

- Es necesario definir índices debido a la gran cantidad de información que se ha cargado.

Actualización. En este paso se fijará con qué frecuencia se volverán a cargar datos en el datawarehouse. Recargar datos es lo que se llama carga incremental. Existen dos maneras de recargar datos:

- Cargar nuevamente todos los datos de las fuentes independientemente de cuáles se hayan modificado.

- Cargar solamente *Deltas*, esto es, cargar registros nuevos o registros que contengan valores que han sido modificados desde la última carga, esto se controla mediante el valor de un campo timestamp de los archivos. Si los archivos no poseen timestamp habría que extraer la totalidad de datos de la fuente e irlos comparando con los almacenados en el datawarehouse, lo cual sería notablemente más costoso.

Ambas soluciones son válidas, pero la segunda es más eficiente debido a la gran cantidad de datos que se manejan.

Las herramientas ETL son clave en proyectos Business Intelligence, la salida que devuelva la herramienta BI depende directamente de los datos de entrada que se dispongan, tiene una importancia crítica que estos datos pasen los subprocesos mencionados.

2.3.3 Datawarehouse

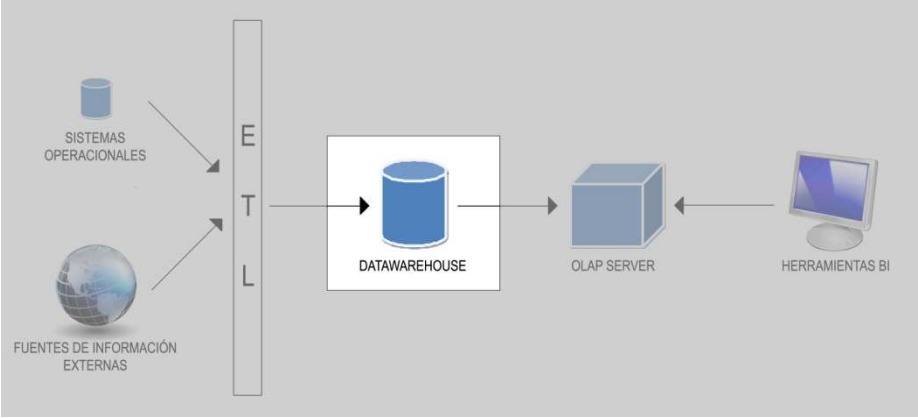


Ilustración 4. Arquitectura-Datawarehouse

Como se viene diciendo en la explicación de los procesos anteriores, una vez listos los datos, se migrarán al datawarehouse. En esta sección va a explicarse en qué consiste un datawarehouse y por qué razón deben trasladarse los datos de los sistemas operacionales a este tipo de estructura.

Se empezará explicando los porqués: El objetivo de una herramienta BI es al final dar soporte para la toma de decisiones, como se ha expuesto varías veces a lo largo del presente documento, para tomar dichas decisiones el usuario se basará en la información, pero la información que contienen los sistemas operacionales hace referencia a transacciones diarias, lo cual no es lo suficientemente óptimo para basarse en ella a la hora de tomar una decisión, ya que al usuario le sería de gran ayuda contar con datos históricos, datos totales, etc.

Si no se construyera un datawarehouse y se recuperara la información directamente de los sistemas operacionales, para analizarla a través de una hoja de cálculo por ejemplo, aparecerían problemas surgidos de los errores y del tiempo la mayoría de ellos:

- Pueden cometerse errores al introducir la información de las diferentes fuentes.
- El proceso de extraer la información de las fuentes operacionales y la introducción en la hoja de cálculo conllevaría mucho tiempo.
- Si se necesitara posteriormente información más detallada, habría que repetir el proceso.
- Habría que repetir el proceso también cuando los datos cambiaran.

Al almacenar los datos disponibles en los sistemas operacionales en el datawarehouse contaremos con la información consistente, integrada y que incluye datos históricos. De esta manera analizar la información existente resultará mucho más sencillo. Con la totalidad de la información sostenida en un solo sitio, la herramienta BI tendrá que consultar un único almacén, en el que además se encontrarán los datos de alta calidad.

“Un datawarehouse es una colección de información creada para soportar las aplicaciones de toma de decisiones” [J.Watson, 2006]

Un datawarehouse debe estar orientado a un área, ser integrado, indexado en el tiempo, y no volátil:

- Orientado a un área: La información se clasifica dentro del datawarehouse para atender a las diferentes áreas de interés de la organización o los diferentes problemas de negocio. Diferentes áreas necesitarán información diferente para

llevar a cabo su actividad, realizando esta clasificación se otorgará a cada área de negocio sólo la información que necesita, ignorando el resto, que le resultaría no relevante.

- Integrado: Se integran dentro del datawarehouse los datos obtenidos de las diferentes fuentes de información (sistemas operacionales y/o fuentes externas). La integración se lleva a cabo estableciendo medidas uniformes, y convenciones para nombrar a los datos y para adoptar claves.

- Indexado en el tiempo: En el datawarehouse la información se almacena conforme a diferentes momentos del tiempo, de manera que la información almacenada son fotografías de los distintos periodos. Almacena la información y es gracias a ello por lo que podemos realizar consultas que tengan en cuenta la evolución histórica de los datos.

- No volátil: La información no se modifica, como pasaría en los sistemas transaccionales, lo que significa que los usuarios no la mantienen. Una vez la información esté en el datawarehouse ésta no variará, sólo se incrementará.

El datawarehouse es entonces, una gran base de datos corporativa, pero a menudo se hace necesario su división en unidades más pequeñas, los datamarts. Esto es útil por ejemplo en el caso de grandes organizaciones globalizadas, resultaría muy ineficiente que las diferentes sedes tuvieran que acceder al datawarehouse cada vez que necesitaran acceder a información corporativa. Pero no se emplean solo en ese caso, los datamart pueden ser utilizados para los diferentes departamentos dentro de la empresa, o incluso para los diferentes grupos de trabajo. La separación de los datos se puede hacer con el criterio que la organización quiera, según áreas geográficas, departamentos, periodos de tiempo, modelos de negocio, etc.

Ya que están dirigidos a un número menor de usuarios, los datamart almacenan sólo una parte de la información total que estaría almacenada en el datawarehouse, por ello son normalmente más pequeños.

Existen varias formas de integrar los datamarts, y dependen de las necesidades de almacenamiento y de la preferencia de las organizaciones.

Algunas ignoran el concepto de datawarehouse e implementan datamarts, cargando en ellos los datos de las fuentes de información, como se ilustra en la figura 5. Se dice entonces que los **datamart** son **independientes.**

La manera de implementar esta solución es construir un datamart e ir construyendo tantos como se necesiten por área de negocio. Esta solución es sencilla de implementar y no requiere mucho presupuesto, pero conlleva el problema de que podrían existir inconsistencias entre los diferentes datamart, y la información corporativa se contradiría.

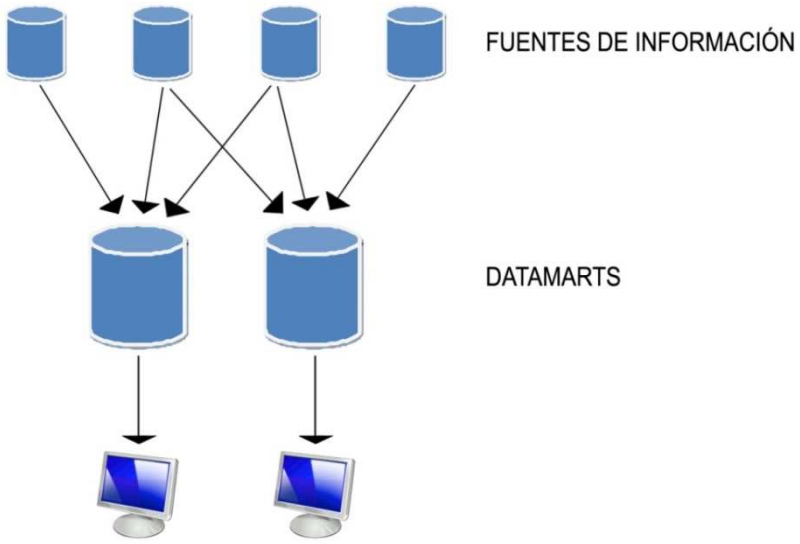


Ilustración 5. Datamarts independientes

Otras organizaciones optan por utilizar los datamarts como estructuras de apoyo que completan al datawarehouse, y que se nutren de los datos de éste, moviendo los datos desde el datawarehouse a los diferentes datamarts, como muestra la figura 6.

La manera de implementar esta solución consiste en construir un datawarehouse, del que se construirá un datamart dependiente que albergará parte de la información del datawarehouse, y se continuaran construyendo todos los datamarts necesarios.

Esta solución aporta un datawarehouse integrado (los datamart no tendrán inconsistencias entre sí) y escalable (se podrán ampliar tantos datamarts como se requieran). Como contrapartida es una solución más costosa y laboriosa de implementar que la solución anterior.

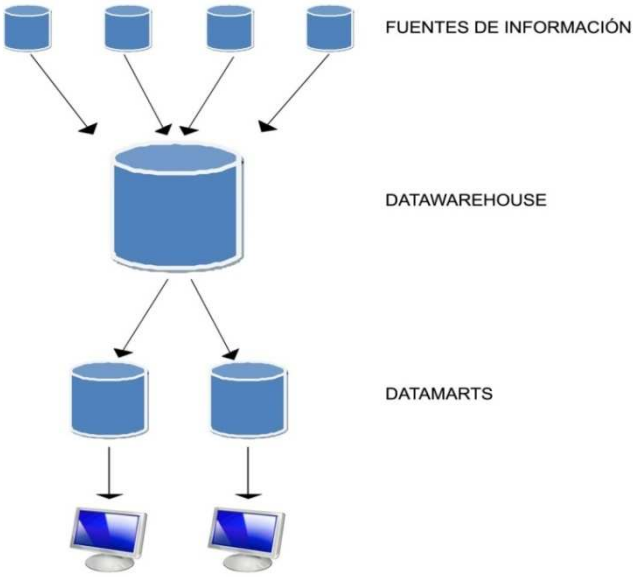


Ilustración 6. Datamarts dependientes

Existe una tercera opción que es una solución híbrida de las dos anteriores, mostrada en la figura 7, consistente en implementar los diferentes datamarts que sean necesarios en un primer paso, que cogerán los datos por lo tanto de las fuentes de información y posteriormente replicar estos datos en un datawarehouse. Sería como la solución de **datamarts dependientes**, pero la copia se realizaría en el sentido contrario: de los datamarts al datawarehouse.

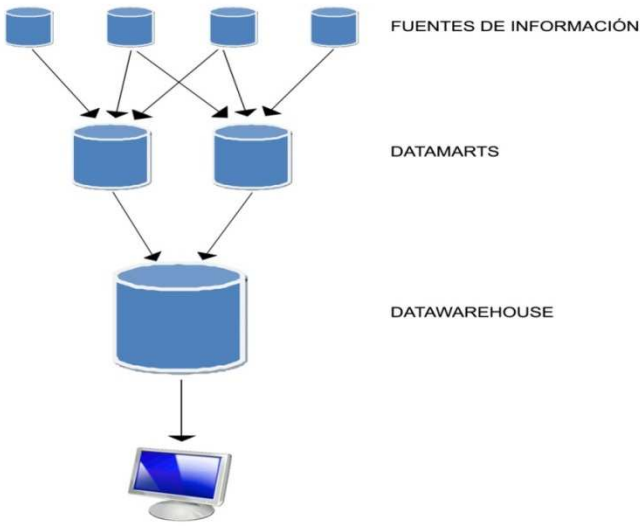


Ilustración 7. Solución híbrida

Las tres estrategias pueden ser válidas si se implementan correctamente y se controla la inconsistencia de datos.

Un componente esencial de los datawarehouse es el **metadata**, es un depósito que almacena metadatos (datos sobre los propios datos). El metadato almacena información sobre qué tablas existen, qué columnas posee cada una de las tablas y qué tipo de datos se pueden almacenar. Esta información es necesaria para los programas que manejarán los datos. Construir el metadata supone por lo tanto definir cada tabla y qué atributos permite, lo que conlleva a nivel empresarial a definir los conceptos de negocio que se tengan y llegar a un consenso sobre ellos.

Los datawarehouse son de carácter no volátil, como se ha visto, por lo que se recargan/actualizan cada cierto tiempo acordado por la organización, no se actualizan en tiempo real, de hacerlo perderían el sentido que tienen, ya que la información cambiaría continuamente y sobre una información cambiante es difícil encontrar un criterio para adoptar una decisión.

De necesitar la organización un almacén que actualizase la información en tiempo real se podría implementar un ODS, muy confundido con los datawarehouse. Pero estos últimos tienen un carácter volátil.

Un ODS (Operational Data Store) es un almacén que consolida datos que provienen de diferentes fuentes, facilitando así el acceso a dichos datos. La principal diferencia que tiene con un datawarehouse es que la información que almacena es volátil, esto es, se actualiza en intervalos de tiempo menores a los que lo hace el datawarehouse.

La razón que podría llevar a implementar un ODS sería por lo tanto la necesidad de una mayor actualización de los datos, ya que en los demás aspectos no es diferente a un datawarehouse. En ocasiones los ODS son tomados como una fuente de datos más para la construcción de un datawarehouse.

Los usuarios finales, se apoyarán en la información contenida en los datawarehouse para tomar decisiones, por lo que es necesario garantizar la disponibilidad y el rendimiento de la estructura, además de contar con copias de backup y permitir la recuperación física en caliente (sin desmontar las bases de datos).

Como resumen, se enunciarán las principales ventajas que aporta un datawarehouse.

- Constituye una herramienta para la toma de decisiones en cualquier área funcional, basándose en información integrada y global del negocio.

- Proporciona un entorno que permitirá encontrar nuevos conocimientos y generar valor a la organización

- Permite observar los datos del pasado, identificando patrones y tendencias que ayudarán a predecir valores del futuro.

- Supone una optimización tecnológica que da unos retornos de inversión muy altos.

2.3.4 OLAP Server.

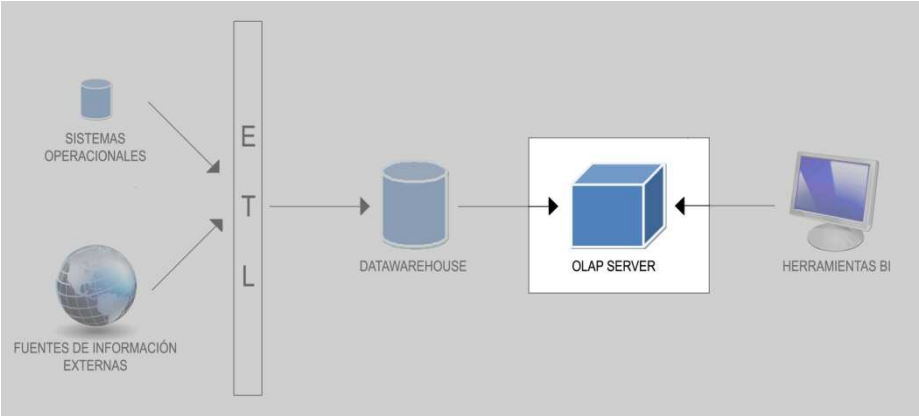


Ilustración 8. Arquitectura-OLAP Server

OLAP del inglés On-Line Analytical Processing es la tecnología más extendida para la recuperación de datos del datawarehouse.

OLAP es definido como el análisis multidimensional e interactivo de la información de negocios a escala empresarial.

Desglosando la definición:

- Es interactivo porque los usuarios se moverán a través de la información, realizarán consultas de la misma, como si se encontraran navegando en ella. Así el usuario se mueve de una perspectiva de su negocio a otra teniendo disponible toda la información de la empresa.

- Es una tecnología multidimensional porque permite extraer la información desde diferentes perspectivas, que son las dimensiones, como se verá más adelante.

- Es a escala empresarial, ya que dispone de toda la información existente en la organización.

Los usuarios trabajarán de manera intuitiva con la información, sin necesidad de saber cómo es el formato físico de los datos, conocer el lenguaje SQL, los nombres de las tablas, la arquitectura OLAP o demás conocimientos técnicos.

El procesamiento OLAP tiene bastantes diferencias de base sobre el procesamiento OLTP.

OLTP (On-Line Transaction Processing) es el procesamiento de la información que emplean los sistemas operacionales, el llamado procesamiento operacional. Se centra en procesar las actividades diarias de la organización, como pueden ser altas, bajas, modificaciones, etc. Las aplicaciones con un procesamiento OLTP están orientadas a la carga de datos.

Por el contrario OLAP se emplea para las actividades de la organización que no son meramente de carga de datos, sino más bien de toma de decisiones sobre el propio negocio (que es al final el objetivo de la herramienta BI), como son la planificación estratégica, de recursos, iniciativas de mercado, etc.

OLAP nació porque no todo son transacciones. ¿Qué sucede si se quiere analizar tendencias?, como por ejemplo:

- Tratamientos más utilizados por los expertos para una determinada patología, etc.

- Proyectos totales llevados a cabo por meses, países o por tipo de patología que afectaba a la edificación, etc.

- Comunidad que realizó más proyectos el mes pasado, etc.

OLTP sostiene una evidente complejidad para visualizar los datos y formular consultas. Con OLAP pueden realizarse consultas y visualización de datos de acuerdo

al modelo de datos multidimensional y se nos ofrecen las respuestas en muy pocos segundos.

Las características técnicas que diferencian a OLAP de OLTP son las siguientes:

- En una aplicación OLTP los datos se actualizan con la misma frecuencia que son leídos, mientras que en una aplicación OLAP se actualizarán conforme lo haga el datawarehouse subyacente.

- El tiempo de respuesta en una aplicación OLTP es más crítico que en una aplicación OLAP ya que maneja datos de la actividad diaria de la empresa, y pudiera tratarse de una transacción importante para un cliente, por ejemplo.

- Las aplicaciones OLTP son usadas por un gran número de usuarios dentro de la organización debido a la información que tratan, por lo que deben soportar un alto nivel de concurrencia, mientras las que una aplicación OLAP en principio está destinada a un número menor de usuarios que serán los tomadores de decisiones de la organización.

- La frecuencia con la que se trabaja con una aplicación OLTP es bastante constante, ya que realiza operaciones diarias de la empresa, mientras un OLAP puede utilizarse de manera diaria en algunos periodos y mensual en otros.

- Y la diferencia más importante es que las consultas que se realizan sobre OLTP son de carácter más sencillo y sobre una menor cantidad de datos que las realizadas sobre OLAP, que suelen ser consultas complejas que conllevan el análisis de gran cantidad de datos.

En la tabla 2 se resumen las principales diferencias.

	OLTP	OLAP
Usuario típico	Operacional	Analista
Función	Almacenamiento/consulta de datos	Análisis
Registros accedidos	Decenas	Millones
Unidad de procesamiento	Transacción simple	Agregación
Redundancia de datos	Causa ineficiencias	Clave para procesamiento eficiente
Modelo de datos	Relacional	Multidimensional

Tabla 2. Comparación OLTP-OLAP

Como se ha dicho, OLAP permite utilizar la información desde diferentes perspectivas, a estas perspectivas se las llama dimensiones porque conforman las diferentes dimensiones del cubo. El cubo es la estructura multidimensional que utiliza OLAP para albergar la información, que gráficamente representado es como se ilustra en la figura 9.

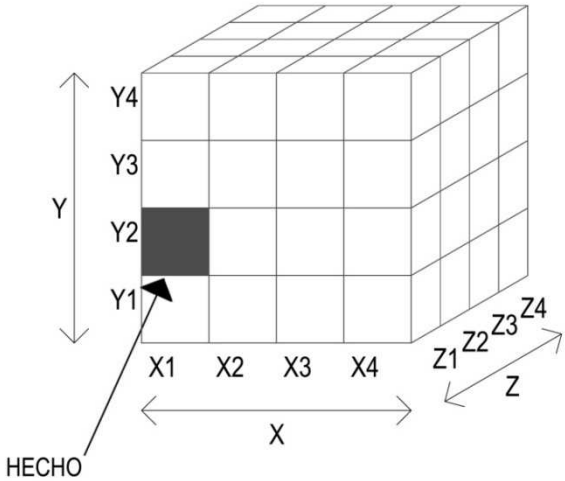


Ilustración 9. Cubo genérico

Los ejes ‘x’, ’y’ y ‘z’ son las dimensiones del cubo y los valores x_1, x_2,…,x_n ; y_1,y_2, …,y_n ; z_1,z_2,…,z_n son las medidas sobre las que se analiza la información.

Las dimensiones son los elementos sobre los que interesa analizar tendencias, comportamientos, etc. Cada organización definirá las dimensiones que le parezcan de mayor interés para el análisis.

El cubo queda dividido así en una matriz con múltiples entradas, que permiten hacer diferentes combinaciones de consulta, visualizando los datos desde las diferentes perspectivas del cubo.

El contenido de cada posición de la matriz, de cada cubo individual, se llama “hechos”. Las soluciones OLAP actuales permiten que cada cubo contenga varios hechos.

El cubo es una estructura independiente a los sistemas operacionales de la empresa, y al datawarehouse. Es montado con posterioridad y su función es facilitar la consulta de la información.

Un ejemplo sencillo de cubo es el que se ilustra en la figura 10.

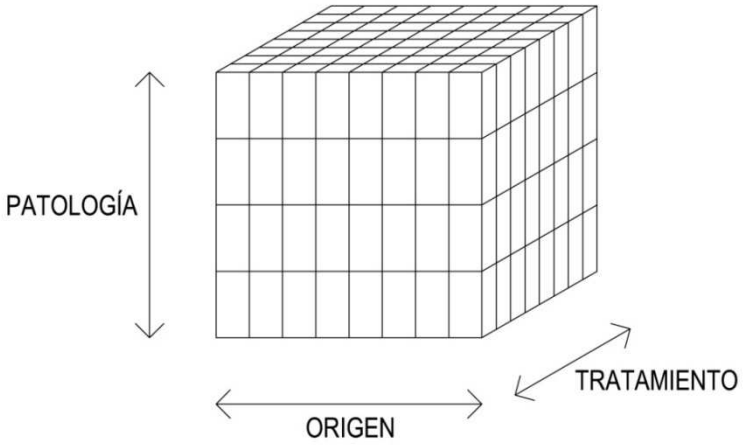


Ilustración 10. Cubo ejemplo: patologías arquitectónicas

En este cubo se tienen los hechos de cada patología registrada, con el correspondiente tratamiento y origen.

Sobre el cubo pueden realizarse diferentes operaciones multidimensionales, que se clasifican en tres grupos.

Grupo 1.Operaciones de selección.

- Slice. Consiste en seleccionar una 'rebanada' del cubo, creando así un subcubo menor. Por ejemplo si se seleccionan los datos de productos vendidos por país sólo para la temporada de otoño, se obtendrá solo un subconjunto de los datos del cubo completo, como se muestra en la figura 11.

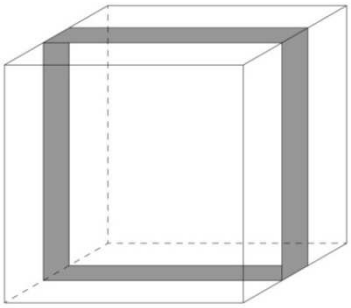


Ilustración 11. Operación slice

- Dice. Tomar secciones del cubo en función a un rango de valores. Si al sub-cubo creado antes con la operación *slice*, se le restringe indicando que se quieren obtener los valores de conjunto correspondientes al producto cuyo valor haya superado las 3000 unidades vendidas. Al hacer esta operación se seleccionan solo los hechos que cumplan con el rango especificado, en esto consiste la operación *dice*.

- Pivoting. Esta operación consiste en rotar el cubo para poder observarlo en diferentes planos. Rotar el cubo significa cambiar el orden de las dimensiones del mismo.

  En la figura 12 se ilustran las tres posibles rotaciones del cubo para el ejemplo utilizado.

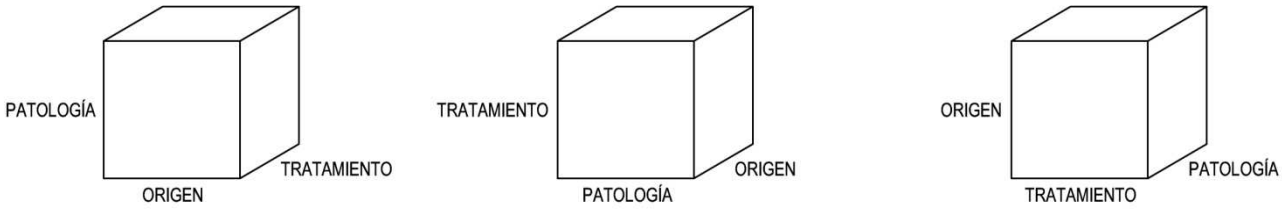


Ilustración 12. Rotación del cubo

Grupo2. Operaciones de Agregación.

Mediante las operaciones de agregación nos desplazamos a través de las jerarquías y niveles de las dimensiones.

- Drill-down. Consiste en bajar un nivel en la jerarquía. Así si por ejemplo se están observando los productos vendidos anualmente en un país, al realizar la operación drill-down se segregan esas ventas por temporada. Dicho en otras palabras, al hacer drill-down, bajamos en el nivel de detalle de los resultados buscados.

- Drill-up. Operación contraria a la anterior. Al hacer drill-up, nos abstraemos más sobre el nivel de detalle anterior a realizar la operación.

  Cada vez que se hace drill-up o drill-down se está creando una nueva consulta de tipo join.

  Hay que decir también, que en las aplicaciones BI suelen realizarse consultas complejas, que son el tipo de consultas en las que mediante múltiples consultas consecutivas, se hace una selección progresiva de lo que se está buscando.

- Roll-up. Representa el máximo nivel de agregación, sirve para el cálculo de totales. Así, podría interesarnos realizar esta operación para averiguar el total de tratamientos por patología y origen de la misma. La imagen 13 representa está operación.

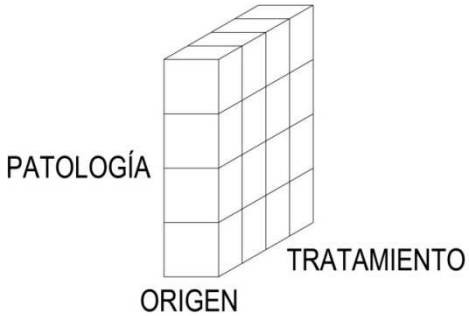


Ilustración 13. Operación roll-up

Grupo 3. Operaciones de Relación.

Las operaciones de relación son las que permiten acceder a otros datos.

- Drill-across. Se accede a datos que están en otros cubos.

- Drill-throught. Mediante esta operación se accede a datos que se encuentran en el datawarehouse o en un sistema operacional.

Dependiendo de cómo la estructura OLAP almacene los datos se habla de diferentes tipos de herramientas OLAP:

ROLAP

Del inglés, Relational OLAP. Son herramientas OLAP que acceden a la información de bases de datos relacionales. Una herramienta ROLAP requiere de un datawarehouse sobre el que acceder para recuperar los datos. Las herramientas ROLAP realizan mediante consultas SQL un mapeo de los datos almacenados en el datawarehouse a un modelo multidimensional.

Capturar los datos de una base de datos relacional tiene ventajas, y por eso en algunas ocasiones resulta la solución más eficiente, las siguientes son algunas de las ventajas de su utilización:

- Son escalables a conjuntos grandes de datos, lo que permite aumentar el volumen de datos sin tener que recargar los ya existentes, como ocurre en MOLAP.

- Pueden manejarse volúmenes de datos grandes, cuando se tiene un gran datawarehouse, esta ventaja es primordial.

- Los empleados de sistemas suelen estar más habituados al uso de bases de datos relacionales.

MOLAP

Del inglés, Multidimensional OLAP. Son herramientas que utilizan almacenamiento multidimensional, es decir acceden a las llamadas bases de datos multidimensionales (MDDB).

Son estructuras indexadas que permiten un excelente tiempo de acceso a cualquier elemento.

La principal ventaja de las MDDB es la pre-consolidación de los datos. Cuando se realiza una consulta de un total, éste se calcula rápidamente y luego puede almacenarse el resultado en la propia estructura para un uso posterior, reduciendo así el tiempo de respuesta de la próxima vez que sea consultado. Como pre-consolidar los datos requiere mucho espacio de almacenamiento, la alternativa es almacenar los totales más utilizados.

Además las MDDB cuentan con mecanismos para el trato de datos dispersos. El concepto de dato disperso surge del hecho de que es posible que no todas las celdas de la estructura contengan hechos. Con el ejemplo explicado anteriormente del diagnóstico de patologías, puede darse el caso de que para un determinada patología y un determinado origen no existan ningún tratamiento asociado, en este caso se tendrá una celda del cubo vacía. Esto resulta ineficiente y por lo tanto las herramientas MOLAP tienen algoritmos para evitar guardar valores nulos.

La estructura MOLAP puede organizarse de dos maneras diferentes:

- Puede crearse un hipercubo, que consiste en un cubo de gran tamaño que almacena todos los datos. Esta es la manera más simple y más eficiente en cuanto a tiempo de respuesta cuando se lanza una consulta sobre el cubo, ya que se tiene la información más centralizada.

- Pudieran crearse diferentes cubos en los que repartir los datos, normalmente datos relacionados. Esta organización mejora el tiempo de respuesta de las consultas, si éstas solo requieren acceso a un solo cubo, pero desmejora el tiempo de respuesta, respecto a la solución del hipercubo, en caso contrario.

Algunos de los inconvenientes que presenta MOLAP es que el análisis se limita prácticamente a datos totalizados, debido a que pueden albergar menos información de las bases de datos relacionales.

Y, por otro lado, un modelo multidimensional no se adapta bien a los cambios en la organización, ya que si se quieren cambiar las dimensiones del cubo, se debe crear un nuevo cubo.

La ventaja principal respecto a ROLAP, es la trivialidad de los tiempos de respuesta, ya que el almacenamiento se encuentra optimizado para análisis multidimensional.

HOLAP

Del inglés, Hybrid OLAP. Las herramientas HOLAP son una combinación de las MOLAP y las OLAP, la idea es quedarse con lo mejor de cada una de las soluciones.

Es una herramienta que almacena datos totalizados en una MDDB, pero que accede a una base de datos relacional para obtener datos más detallados (como se ha dicho, la base de datos relacional guardará mayor cantidad de datos, imposible de almacenar en una MDDB, luego se accederá a ella para datos no guardados por problemas de espacio, en esta última).

2.4 Modelo dimensional

En el caso del presente PFC se cuenta con una base de datos sobre la que se tiene el interés de analizar. De no ser así, en proyectos que partan sin un modelo de datos, se deben pensar tres cosas para construir el modelo:

1. Cuál es el modelo de negocio de la organización.
2. Cuáles son los indicadores de negocio.
3. Definir el modelo de datos. ¿De qué información se dispone? Al definir este modelo hay que decidir qué hechos se desean analizar y bajo qué dimensiones.

Lo vemos a continuación con más detalle.

1. Modelo de negocio

Un modelo de negocio es una representación de la realidad a partir de la cual se puede analizar lo que sucede en la realidad. Los modelos sirven para ayudar a predecir en qué forma afectarían cambios que se introdujeran, pudiendo comprender qué parte del resultado se debe a qué factores.

Si no se conoce el modelo de negocio, no puede estarse en disposición de analizar la información, ¿Con respecto a qué se analizaría? No puede analizarse un entorno que no se conoce. Para construirlo se debe documentar, probar y desarrollar las teorías existentes de cómo funciona el negocio. Existen varías tecnologías que pueden aplicarse para definir el modelo, como puede ser el análisis de procesos, análisis de ratios, EFQM, etc.

En resumen construir el modelo de negocio es documentar la manera en que funciona el mismo.

2. Indicadores de negocio

Una vez conocido el modelo de negocio, el segundo paso es conocer cuáles son los indicadores de ese negocio. Los indicadores de negocio o KPI (de sus siglas en inglés, Key Performance Indicator) son variables cuantificables relacionadas con un objetivo. Son los factores, definidos por la propia organización, que se consideran clave para la

ejecución del negocio de la empresa. Estas variables ofrecen una perspectiva de la variación por exceso o por defecto de un resultado respecto de un objetivo marcado, sirven por lo tanto para medir el nivel de mejora de las actividades fundamentales de la organización.

Los KPI pueden representarse como un ratio, de manera que pueda observarse instantáneamente la situación de esa variable con respecto a la planificación, sin ser necesario entrar en detalle sobre las diferentes métricas que pueden componer el KPI.

Es necesario conocer los KPI de la organización para conocer el tipo de decisiones que se toman en dicha organización, para poder averiguar qué datos son necesarios para apoyar esas decisiones. Para conocer los KPI se debe implicar a los futuros usuarios de la herramienta BI, porque son los que pueden determinar la importancia de determinados procesos o datos de negocio que les permiten tomar decisiones y conocer el rendimiento.

Para esto pueden realizarse a estos usuarios una serie de entrevistas para poder educir el conocimiento de los KPI´s. De estas entrevistas deben educirse dos conceptos. El primero es comprender la medida, la dimensión y el objetivo que representa cada KPI. Una vez comprendidos los KPI que responden a sus necesidades, lo segundo es averiguar si los usuarios van a querer hacer un análisis posterior o directamente tomar una decisión o acción basándose en el resultado numérico que devuelva los indicadores.

Las preguntas que pueden realizarse en las entrevistas se pueden organizar en cuatro bloques principales, que responden a diferentes cuestiones, que son las que se busca responder.

- Bloque 1. ¿Qué representa el KPI?

  Las preguntas que se realicen de este bloque estarán dirigidas a comprender qué información esperan los usuarios obtener del indicador de negocio y qué uso le darán a esta información.

- Bloque 2. ¿A qué personas está destinado el KPI?

  En este bloque se pretende conocer el rol de personas para las que el resultado del KPI aportará valor. Sabiendo esto puede conocerse más el

indicador, si se relaciona con la función que desarrollan estas personas dentro de la organización.

- Bloque 3. ¿Por qué es importante cada KPI?

  Conocer la importancia de cada indicador ayudará a construir la aplicación y a tener una visión general de los procesos de negocio de la organización, porque se conocerá la información más relevante de entre la demandada y podrá resaltarse en la aplicación sobre el resto o podrá por ejemplo ser relacionada en mayor medida con otras clases, se 'jugará' más con esa variable.

- Bloque 4. ¿Fuentes de datos?

  Por cada indicador deben conocerse los datos que lo soportan y dónde se encuentran, es posible que sea necesario agregar datos de varías fuentes de información para conformar un KPI.

Concluidas las entrevistas a los usuarios, hay una serie de conceptos que deben conocerse de cada indicador de negocio: su nombre, la forma en que se calcula, y el objetivo que tiene, es decir, los valores que se consideran óptimos para ese indicador.

3. Modelo dimensional

El último de los pasos, y el objetivo del proceso descrito, es la realización del modelo dimensional. Para construirlo se debe determinar cuál es la información relevante, interesante para analizar el modelo de negocio, es decir, debe definirse el modelo de datos.

Esta información puede organizarse en un modelo relacional. El modelo relacional es un tipo de modelo de datos en el que éstos se organizan en tablas definidas por su clave primaria y el resto de sus atributos. Donde la relación entre las tablas se realiza mediante el uso de claves externas FK (de las siglas en inglés, Foreign Key) y gráficamente mediante la unión de esas tablas con flechas.

Los almacenes de información de las organizaciones suelen estar constituidos en bases de datos relacionales. Una base de datos relacional es una opción válida como entrada a un sistema BI. No obstante si se necesita mayor nivel de análisis lo

recomendable es organizar la información siguiendo un modelo de datos dimensional, porque deja observar por ejemplo cómo el negocio se ha comportado a lo largo del tiempo, cómo se comporta en el presente y cómo se estima se comportará en el futuro, y porque permite moverse con mayor agilidad, buscando específicamente entre la información que se necesite, lo que deja entender más claramente qué está sucediendo con un indicador y por qué.

Este tipo de modelo cumple las mismas normas que un E/R pero tiene diferente estructura, que en modelo dimensional se llama diagrama en estrella. Los elementos que tiene el **diagrama en estrella** son dos: las tablas de hechos y las tablas de dimensiones, que se explicarán a continuación, para entender en qué consiste el modelo dimensional.

Tablas de Hechos.

Las tablas de hechos representan aquellos conceptos del dominio que quieren analizarse. Estas tablas están compuestas por los datos que caracterizan al concepto en sí, que suelen ser datos numéricos y aditivos (hecho que es importante porque suelen realizarse operaciones algebraicas sobre estos campos), y también porque contienen una clave ajena por cada tabla de dimensión, esta es la manera de acceder a las dimensiones.

Para definir las tablas de hechos hay que elegir en que datamart vamos a almacenarla. ¿Y por qué en un datamart y no en el datawarehouse directamente? Porque como se vio en el capítulo de arquitectura anterior, los datamart son creados específicamente como un subconjunto del datawarehouse especializado en unos determinados datos, y se podían dividir según diferentes criterios, por ejemplo, por modelo de negocio. Y el modelo dimensional busca precisamente explotar la información según modelo de negocio.

Además las uniones entre tablas que hay en un diagrama en estrella no serían óptimas fuera del ámbito de los datamarts, porque el mantenimiento sería costosísimo. El problema de actualizar se da entonces en los datamarts, pero al ser los datamarts estructuras que se cargan con un periodo determinado, el esquema estrella puede aplicársele.

El siguiente paso es elegir que dimensiones estarán relacionadas con la tabla de hechos, o dicho de otro modo, bajo que dimensiones o medidas se quiere analizar el concepto que representa la tabla de hechos.

Tabla de dimensiones.

Las tablas de dimensión representan las medidas bajo las que se quiere analizar las tablas de hechos, como se quieren medir los conceptos. Estas tablas no son otra cosa que atributos que permiten reunir los hechos, los conceptos a analizar en función de los valores que adquieran.

Una tabla de dimensión contiene atributos descriptivos que la caracterizan, por lo que todos los atributos estarán relacionados conceptualmente.

Un diagrama en estrella admite una sola tabla por cada dimensión, tal y como muestra la imagen 14. Pueden existir varias tablas de hechos que quieran analizarse desde la misma dimensión, entonces estarán unidas a la misma tabla de dimensión, cuando esto sucede se habla de constelaciones de hechos o se puede llamar también diagrama galaxia.

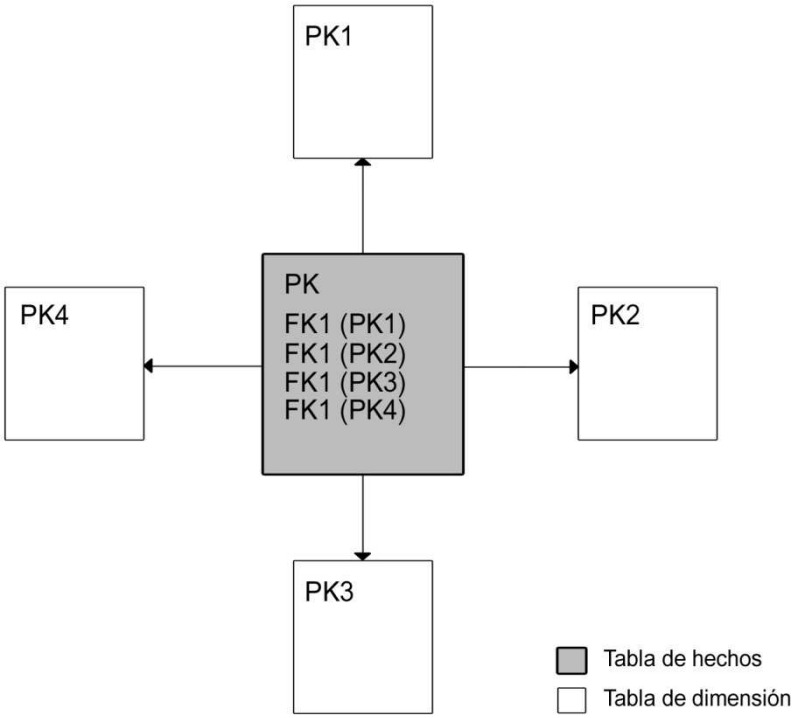


Ilustración 14. Esquema en estrella

Existen también los diagramas llamados en **copo de nieve**. Éstos tienen las mismas características que el explicado diagrama en estrella, pero se encuentra normalizado. Es decir un diagrama copo de nieve es un diagrama estrella normalizado. Para normalizarlo aparecerán relaciones entre las tablas de dimensiones y así evitar las redundancias que aparecían en el modelo estrella.

Un modelo dimensional no es un datawarehouse, las principales diferencias son que el modelo dimensional almacena menor volumen que el datawarehouse y por otro lado que el modelo dimensional tiene infinitas formas de consultarse y el datawarehouse no.

2.5 Selección de la herramienta BI

Se ha expuesto el concepto y arquitectura del BI, pero no se ha entrado a detallar las herramientas y respectivos proveedores que existen actualmente en el mercado, de ello tratará el presente punto.

Tiene una importancia relevante saber escoger la herramienta Business Intelligence adecuada, es posible que un proveedor tenga, por ejemplo, una herramienta muy buena que realice Reporting, pero quizás nuestra necesidad esté más basada en la utilización de cuadros de mando, en cuyo caso pudiera ser más conveniente otra herramienta.

Escoger correctamente ahorrará tiempo y dinero a la organización, ya que elección lleva consigo el pago de la licencia del producto, los costes asociados a la implantación, el coste de posibles cursos de formación a empleados, etc. Si la herramienta es elegida de manera informal, puede que la empresa tenga que invertir en estos costes por duplicado al descubrir que la aplicación contratada no cumple con los requisitos como se esperaba.

Es importante definir por lo tanto cuál es el punto de partida y qué se pretende conseguir de la aplicación BI. Es bueno también anticiparse a necesidades futuras que se puedan tener para que la aplicación se adecue a los requisitos pedidos lo máximo posible, ya no sólo en tiempo presente, también a medio o largo plazo para garantizar una rentabilidad a la inversión realizada.

Se deben contemplar las características de los usuarios que manejarán la herramienta. Existen dos grandes roles de usuario de las herramientas BI los productores de información y los consumidores de información:

- Los productores de información: Se trata del personal del servicio técnico o de usuarios avanzados. Serán los encargados, en la mayoría de los casos, de crear los informes o modelos que posteriormente explotarán los usuarios finales. En definitiva, son usuarios que poseen conocimientos de informática y entienden y pueden manejar la información almacenada.

- Los consumidores de la información: Se trata de los usuarios finales de la aplicación, aquellos que consultarán los informes para la posterior toma de decisiones, en este grupo se encuentra todo el personal que la organización

crea oportuno, la dirección será la encargada de decidir quién tendrá acceso a qué información. En general el grupo lo conforman directivos, responsables y jefes de proyectos.

Diferentes roles necesitarán funcionalidad diferente de la herramienta, como se refleja en la figura 15:

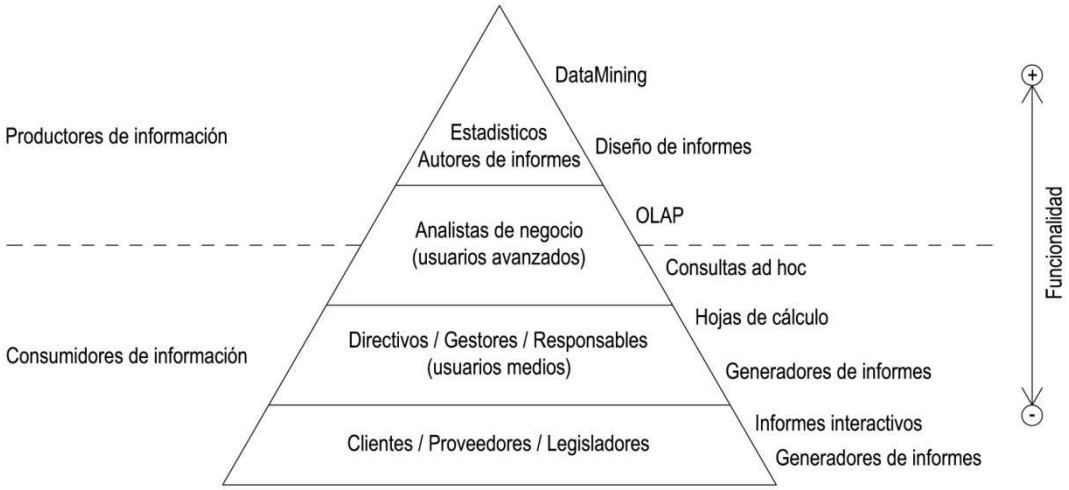


Ilustración 15. Pirámide herramientas-rol

Conocer los roles de los diferentes usuarios de la aplicación ayuda a escoger qué herramienta se adecua mejor a las necesidades de éstos en cada caso. Así por ejemplo un usuario que produce información utilizará en mayor medida una solución de datamining, pero a un usuario tomador de decisiones le será de mayor utilidad herramientas como cuadros de mando, informes interactivos, etc.

Existe una gran variedad de herramientas en el mercado, y el trabajo de elegir una de ellas no es trivial. Si no se aplica un método formal para la elección de la herramienta, es posible que ésta no repercuta los beneficios esperados.

No es conveniente, por lo tanto, menospreciar la importancia de este punto del proyecto BI, y adquirir la herramienta del proveedor que mejor haya realizado la demostración.

Es necesario aplicar un proceso formal de elección de la herramienta, siguiendo una serie de pasos que ayudarán a elegir la más conveniente a nuestro caso, los pasos son los siguientes [W. Eckerson et al. 2005]

1. *Construir un comité* que será el encargado de la elección de la herramienta, estará constituido por todos los interesados de los diferentes departamentos, incluido también el personal informático de la organización. No es conveniente asignar la responsabilidad de la elección a un responsable de negocio o del departamento de informática, se deben reunir un comité de responsables, ya que la aplicación interactuará con gran parte de la organización, una vez se encuentre implantado.
2. *Concretar quiénes serán los usuarios y en qué contexto.*

   Se definirán los roles de usuarios que utilizarán la herramienta, y qué información podrá consultar cada uno de estos roles. Esto es importante porque diferentes roles es posible que necesiten diferentes tipos de herramienta, como se ha visto.

3. *Definir los requisitos*.

   Cada herramienta BI suele estar especializada en la realización de una serie de modelos, su elección depende entonces de para qué se necesite. Es necesario analizar cuál es la especialización que cada proveedor nos ofrece.

4. *Definir criterios de selección y sus pesos*.

   Consiste en plantear los diferentes requisitos de los usuarios como criterios y asignar el peso o importancia que tiene ese requisito.

   Éste es un método práctico que ayuda a minimizar la subjetividad y asegurar que las diferentes herramientas se evalúan de forma coherente.

5. *Solicitud de información* a los diferentes proveedores, para resolver cuestiones sobre las diferentes plataformas por las que la organización no se termina de decidir.
6. *Demostraciones*.

   El comité debe reunirse con los diferentes proveedores para que éste le muestre una demo de la aplicación, esta demo será de mayor utilidad si está basado en datos obtenidos de las fuentes de información de la organización, ya que así, el comité entenderá en mayor medida el ejemplo y comparará con las herramientas de competencia con mayor criterio.

7. *Elección*.

   En este momento, y con conclusiones de los pasos anteriores realizados, se decidirá cuál es la herramienta que se ajusta en mayor medida.

Con los criterios definidos en el punto cuatro del proceso, cada herramienta obtendrá una puntuación conforme satisfaga cada uno de los requisitos.

Cada puntuación se multiplicará por su peso, obteniendo un valor para cada requisito. El sumatorio de todos los valores, nos dará la puntuación total que obtiene la herramienta. La herramienta con mayor puntuación sería en un principio la más interesante.

8. *Probar el concepto*.

   En este momento la organización probará la herramienta en su propio entorno, de esta manera se comprobará si aporta lo que se esperaba de ella y nos sirve para generar la información que precisamos. En este punto, si aún se tenía más de un candidato, se toma una decisión por uno de ellos.

Al final de la aplicación de este proceso, se tendrá una decisión sobre la herramienta. Sería conveniente además, que la herramienta elegida cumpliese una serie de características:

Facilidad de uso.

- Para los usuarios finales. La función principal de la herramienta será proveer acceso a la información de la organización, y será utilizada por trabajadores que poco o nada sabrán sobre el origen de los datos, o su definición. Por lo tanto la facilidad en el manejo de la herramienta es un requisito principal que se debería exigir.

  Como se ha visto no todos los usuarios plantearán las mismas necesidades, algunos utilizarán la aplicación para la consulta de información, pero otros requerirán realizar sus propios informes, por lo que se necesitará de una herramienta integrada que permita realizar todas estas funciones, sin necesidad de un cambio de aplicación.

- Para el departamento de informática. El departamento de informática será en muchos casos el encargado de crear los informes. Se pretende que la aplicación agilice el proceso de toma de decisiones, por lo que es esencial que no conlleve mucho tiempo realizar dichos informes, se buscará por lo tanto una herramienta simple y fácil de administrar.

Fácil de usar significa también que la herramienta sea capaz de integrarse con la infraestructura IT existente, resultando compatible con diferentes navegadores, servidores de aplicaciones o servidores WEB.

Escalabilidad.

- Escalabilidad de acceso a la fuente de datos. Dentro de la organización, la información puede encontrarse en diferentes lugares o fuentes. Los usuarios requerirán acceder a esta información, independientemente de dónde se encuentre. La solución BI escogida deberá permitir acceso al mayor número de fuentes posible, permitiendo combinar esta información de diferentes procedencias en un mismo informe.

- Escalabilidad de los usuarios. Una vez implantada la herramienta BI será utilizada por muchas personas dentro de la organización, por lo que deberá ser capaz de soportar la carga de usuarios tanto en términos de respuesta como de fiabilidad.

  Se debe para ello prever el número de usuarios que la utilizarán contemplando que con el tiempo el número de usuarios de la herramienta crecerá, y la herramienta debe poder seguir soportándolo.

Seguridad.

La seguridad es un punto muy importante a considerar. La herramienta BI mostrará información de la empresa, no solo a los empleados sino también a clientes, socios y proveedores. Por lo que la solución BI adoptada debe proporcionar una capa de seguridad central para todas las aplicaciones de acceso a la vez que permita la integración en la infraestructura de seguridad IT existente.

2.5.1 Plataformas BI

2.5.1.1 Magic Cuadrant de Gartner

Se presentarán en este sub-apartado las principales soluciones BI existentes en la actualidad. Las más relevantes se encuentran representadas en el *Magic Cuadrant* (MQ) de Gartner, (ilustraciones 16 y 17).

Ilustración 16. MQ 2010

El *MQ* es una herramienta de investigación desarrollada por Gartner Inc. De acuerdo con Gartner, el Cuadrante Mágico tiene como objetivo proporcionar un análisis cualitativo en un mercado y su dirección, madurez y los participantes, lo que posiblemente permita a una empresa convertirse en un competidor más fuerte en ese mercado.

Gartner no promueve a ningún proveedor, producto o servicio descrito en el Cuadrante Mágico y desaconseja a los usuarios tecnológicos la selección de las empresas estrictamente posicionadas en el cuadrante de Líderes. El Cuadrante Mágico se trata

simplemente de un estudio de las plataformas presentes en el mercado, y debe ser visto como una herramienta de referencia y no como una guía a seguir.

Utilizando una metodología que Gartner nunca divulga, colocan las plataformas en uno de los cuatro cuadrantes:

- Líderes (Leaders)
- Rivales (Challengers)
- Visionarios (Visionaries)
- Jugadores del sector

Líderes

El sello distintivo de los líderes es que se centran en invertir en sus herramientas hasta el punto que lideran el mercado. Como resultado, los líderes pueden convertirse en ejemplos a seguir y dan directrices con sus productos al resto de compañías de cómo debe ir evolucionando el mercado. Son empresas que poseen una gran trayectoria, y disfrutan de una gran visibilidad en el mercado.

Los líderes suelen ser los proveedores de un amplio nicho del mercado, ya que desarrollan los requisitos generales del mismo. Sin embargo, puede que no respondan a necesidades específicas de empresas que necesiten una herramienta más especializada.

Rivales

Los rivales o challengers tienen una gran capacidad para establecerse. Son empresas grandes con buenas herramientas, que rivalizan directamente con las ofrecidas por los leaders.

Las grandes empresas pueden fluctuar entre los proveedores Challengers y los líderes según sus ciclos de producción o según un cambio en las necesidades de la compañía.

Visionarios

Los Visionarios son los proveedores que tienen la capacidad de predecir sobre cómo evolucionará el mercado, pero que sin embargo tienen menos capacidad comprobada de ofrecer servicios en esa visión. En los mercados tempranos, este estado es normal. En los mercados más maduros, puede reflejar una estrategia competitiva para empresas más pequeñas o incluso para empresas grandes que tratan de diferenciarse.

A menudo tratan de introducir nuevas tecnologías, servicios o modelos de negocio, para lo cual necesitan construir una fortaleza financiera, tener servicio y soporte, ventas y canales de distribución. Supone una apuesta en firme por su producto.

El hecho de que los visionarios se lleguen a convertir en challengers o en líderes depende de cómo las empresas acepten la nueva herramienta o tecnología que ofrecen.

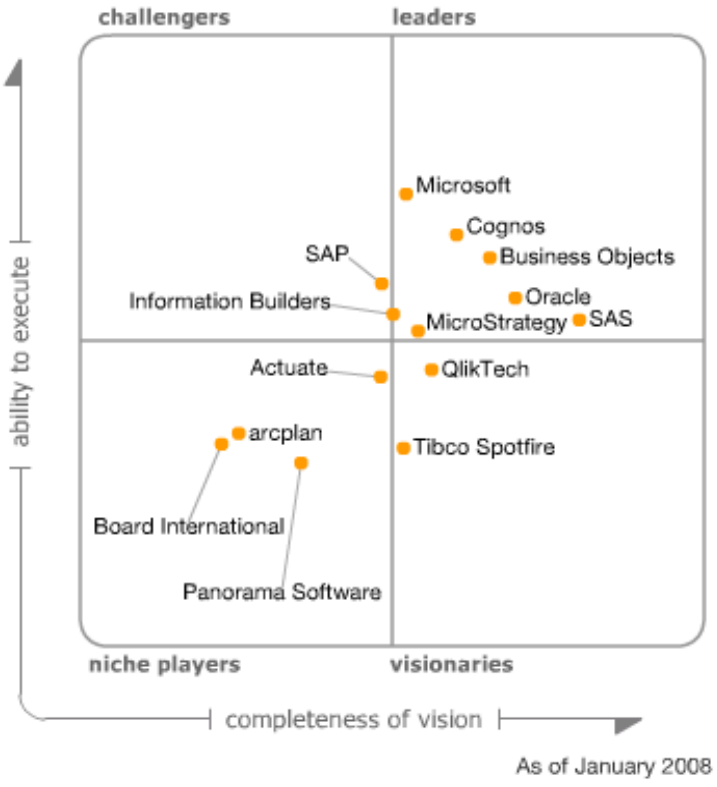


Ilustración 17. MQ 2008

Jugadores del sector

Los niche players o jugadores del sector son proveedores que tienen herramientas que se adaptan solamente a un nicho de mercado, o bien son compañías con una capacidad limitada para innovar o superar a otros proveedores. Esto puede ser porque se centran en una funcionalidad o la región geográfica, o acaban de llegar al mercado.

Las herramientas de estos proveedores pueden tener una funcionalidad bastante amplia pero con escasa aplicación y capacidades de soporte y una lista de clientes relativamente limitada. No tiene aún visión de su producto.

No significa que las herramientas de estos proveedores sean malas, depende de las necesidades de organización, una de estas soluciones puede ser un ajuste perfecto para las necesidades de los usuarios de la empresa cliente. Sin embargo, si va en contra de la dirección del mercado puede ser una elección arriesgada, porque su viabilidad a largo plazo puede verse amenazada.

Cómo interpretar el MQ

Las necesidades y circunstancias de cada empresa deben determinar cómo se utiliza el MQ y no al revés.

Considerar óptimos solamente a los proveedores líderes es una opción arriesgada. Lo óptimo es utilizar el MQ como una guía ante la duda, pero el cliente debe siempre elegir al proveedor que satisfaga sus necesidades en mayor medida.

Por ejemplo, un proveedor en el cuadrante de los niche players podría ofrecer funciones que se adaptan perfectamente a las necesidades de una empresa recién creada y con necesidad de una herramienta BI. Sin embargo un líder pudiera no ofrecer productos que cumpliesen con las especificaciones requeridas, por ejemplo, su oferta puede costar más que los competidores, o que no sea un producto que cumpla completamente con el sector al que pertenece la empresa cliente

El MQ debe utilizarse por lo tanto para limitar la lista de opciones, pero no para basar la decisión solamente en el modelo.

La posición en el MQ brinda información acerca de los proveedores, así por ejemplo si se desea hacer una inversión estratégica en una tecnología, la viabilidad de un proveedor será fundamental, el MQ serviría en este caso para determinar qué proveedores pueden ofrecer la estabilidad necesaria. Se evaluarían en este caso los challengers antes que los visionarios. Por el contrario, si el cliente tiene como requisito obtener una ventaja competitiva mediante la inversión en una tecnología emergente, evaluaría visionarios antes que challengers.

Es así como debe utilizarse el MQ, partiendo siempre de los requerimientos del cliente y no directamente del MQ.

El posicionamiento global de los proveedores en un Cuadrante Mágico ofrece una vista a alto nivel del mercado. En un mercado maduro, una mayoría de los proveedores pueden agruparse en el cuadrante de líderes, porque la mayoría de los proveedores tenderán a ofrecer un conjunto completo de productos y servicios, con una diferenciación escasa.

Por el contrario, en un mercado emergente, la mayoría de los proveedores se agruparán en el cuadrante de los niche players.

2.5.1.2 Productos BI

Se detallan a continuación algunas de las herramientas más importantes de BI de cada cuadrante del Magic Cuadrant. Las descripciones de estos productos comerciales se basan en información del sitio web de cada producto.

Microsoft BI

Tecnologías de SQL Server 2008 para BI

Permite consolidar los datos de diversas fuentes de forma segura.

Proporciona una plataforma de almacenamiento de datos completa y escalable, que permite a las organizaciones construir grandes almacenes de datos, consolidando datos de múltiples sistemas, en una única solución segura y manejable.

Ayuda a obtener información precisa y oportuna a todos los empleados, lo que conduce a una mejor, más rápida, toma de decisiones relevantes.

Capacidades.

- Permite a los usuarios almacenar y compartir documentos.
- Identifica y limpia automáticamente los datos inservibles.
- Implementa un sistema de base de datos relacionales de clase empresarial.
- Combina datos de una amplia variedad de fuentes.
- Crea y maneja informes.
- Proporciona análisis predictivos.
- Reporte de la información como dashboards, scorecards, reporting y con herramientas de colaboración.

Data Warehousing. Almacenar y conectar la información:

Las soluciones Data Warehousins están diseñadas para almacenar y conectar información de forma segura, lo que hace que el análisis, extracción, transformación y carga de los procesos sea más fácil.

Dado que los datos se agregan, se comprimen y se almacenan en un sistema escalable e integrado, las organizaciones pueden reducir los requisitos de almacenamiento a largo plazo (porque no necesitan contar con la información replicada en varios sitios) y generar informes con mayor rapidez.

Los empleados pierden menos tiempo buscando la información correcta y más tiempo en el trabajo de mayor valor y una mejor toma de decisiones.

Análisis Ad-hoc. Analizar los datos rápidamente a través de las dimensiones de negocio:

Las soluciones Microsoft Data Warehousing están diseñadas para almacenar y conectar con los cubos Microsoft OLAP, lo que hace que realizar consultas ad-hoc sea más rápido y sencillo, y que se puedan analizar grandes cantidades de datos complejos desde todos los aspectos del negocio.

Utilizar la funcionalidad de Microsoft OLAP permite mejorar la información de ventas, marketing, informes de gestión, gestión de procesos de negocio, presupuestos y previsiones, informes financieros y áreas similares.

Microsoft DataMining. Profundizar en los conocimientos para deducir:

Las soluciones de Microsoft DataMining permiten ir más allá del simple análisis de datos. Indican tendencias ocultas, problemas o relaciones entre los datos.

Se extraen patrones y las tendencias y los empleados pueden identificar mejor las causas profundas de los problemas y predecir resultados futuros.

Data Cuality. Proteger la calidad de los datos:

Microsoft Business Intelligence protege la calidad de los datos al integrar los mismos de forma segura desde de las múltiples fuentes de datos.

QlikView de QlikTech™

QlikView dispone de herramientas completas, como QlikView Personal Edition, cuya versión actual es la 9.0 (la utilizada para la realización de la aplicación BI de este proyecto). Las herramientas completas incluyen todo lo necesario para realizar una aplicación a medida.

Pero si se desea una herramienta más especializa, QlikView ofrece también diferentes herramientas clasificadas por sector:

- QlikView para bienes de consumo, minoristas y distribución.
- QlikView para servicios financieros.
- QlikView para ciencias biológicas y de la salud.
- QlikView para servicios de infraestructuras.
- QlikView para la Producción.

- y QlikView para el sector público.

Como ésta es la herramienta de desarrollo elegida, se detallará un poco más en el bloque 3, Aplicación BI, del presente documento.

SAP Business Object

Las soluciones SAP BusinessObjects proporcionan una funcionalidad de Business Intelligence completa que permite a los usuarios tomar decisiones fundamentadas y eficaces basadas en datos y análisis sólidos, dando acceso a la información a todo usuario que lo necesite y en cualquier momento, con una dependencia mínima de otros recursos y de los desarrolladores de TI.

Los usuarios pueden además darle formato a la información, analizarla, examinarla y compartirla. Las soluciones de BI SAP BusinessObjects permiten lo siguiente:

- Análisis avanzados
- Cuadros de mando y visualización
- Infraestructura de información
- Consultas, informes y análisis
- Generación de informes
- Búsqueda y navegación

Los productos comerciales Business Objects son:

SAP BusinessObjects XI 3.0

Que proporciona:

- Una plataforma de BI para entornos heterogéneos, de modo que las empresas pueden ofrecer una visión completa y fiable de todos los activos de información.
- Una interfaz fácil de usar para usuarios de negocios y analistas.

- Flexibilidad de despliegue de BI real, lo cual permite al departamento de IT ofrecer y mantener las aplicaciones de BI para más usuarios en toda la empresa.

SAP BusinessObjects Data Services

El software SAP BusinessObjects Data Services es una tecnología que reúne en un solo producto avanzadas soluciones de calidad e integridad de datos, para ayudar a las organizaciónes a trasladar, transformar y mejorar sus datos.

SAP BusinessObjects Explorer

SAP BusinessObjects Explorer une la simplicidad y rapidez de las funciones de búsqueda con la confianza y el poder de análisis de Business Intelligence. Con esta solución, los usuarios utilizan el método tradicional de búsqueda por palabras clave para localizar la información y explorar los datos directamente.

SAP Business Edge BI paquete estándar

SAP BusinessObjects Edge, paquete estándar, ofrece BI estándar completa a las organizaciones medianas. Este software proporciona una experiencia de BI intuitiva independientemente del lugar y del modo en que se trabaje, permite utilizarla desde dispositivos móviles y cuadros de mando hasta integración con Microsoft Office.

SAP BusinessObjects Edge BI, con integración de datos

Con la versión de SAP BusinessObjects Edge BI que incluye integración de datos, se puede integrar datos de fuentes dispares, así como depurar y verificar la calidad de los mismos, lo que proporciona una base de datos segura y suministra a sus usuarios información precisa, fiable y oportuna.

Oracle Business Intelligence

Oracle BI Suite Enterprise Edition (EE)

Oracle BI Suite Enterprise Edition (EE) es una plataforma de BI completa que ofrece una base para crear soluciones BI empresariales desde fuentes de datos heterogéneas para la distribución de datos, con sistemas Oracle y que no son de Oracle. Oracle BI EE está diseñado para un uso extensivo, con nuevos niveles de uso y alcance con el fin de brindar a un público más amplio conocimientos integrales y puntuales. Oracle BI EE también representa la base tecnológica para la inteligencia de negocios en las aplicaciones Oracle Fusion.

Oracle BI Standard Edition (SE)

Oracle BI Standard Edition (SE), que incluye Oracle Discoverer, está optimizado para trabajar con los datos y las aplicaciones Oracle, y ofrecer análisis e inteligencia avanzados al menor costo.

Oracle BI Publisher

Oracle BI Publisher, también denominado Oracle XML Publisher, ofrece una solución eficiente y escalable para informes y publicaciones, disponible para entornos complejos y distribuidos. Disponible con Oracle BI Suite EE o como solución independiente. Posee una arquitectura central para generar y proporcionar información a los empleados, clientes y socios comerciales, tanto de manera segura como en el formato adecuado.

Oracle Real-Time Decision

El software de Oracle Real-Time Decision combina los requerimientos comerciales y de información del usuario para hacer una buena recomendación en cada decisión operacional al adaptarse de manera inteligente la información en cambio constante.

Targit BI Suite

Target BI Suite proporciona cuadros de mando, análisis e informes. Esta herramienta en detalle está compuesta de los siguientes sub-productos:

Targit Analysis.

TARGIT análisis se utiliza para mostrar y analizar datos de la empresa. El cliente contiene todas las aplicaciones necesarias para generar análisis, dashboards y scorecards.

Los análisis se basan en la información almacenada en el datawarehouse de la empresa. El análisis establece una vista completa de las áreas deseadas y ofrece facilidades para analizar cualquier aspecto, a través de todas las áreas de negocio de la compañía.

Targit Net.

Targit Net es la edición web de Targit. Permite estar en contacto con los datos empresariales en cualquier lugar y a cualquier hora desde cualquier ordenador con conexión a internet.

Targit Power.

Targit Power, realiza procesamiento de datos y análisis numéricos complejos. La combinación de análisis e informes proporciona las facilidades apropiadas para el análisis de datos integrado.

Targit Desktop.

Con el cliente de escritorio pueden verse objetos, gráficos y medidores directamente desde el escritorio de Windows. También puede hacerse clic en los objetos para ver una vista previa del análisis en Targit, y de esta manera se aconseja al usuario de cuándo debe iniciar sesión en Targit para analizar posibles desviaciones críticas. La función se basa en el concepto de ofrecer a los usuarios "tiempo real" en el conocimiento de los datos importantes.

Targit ANTserver.

TARGIT ANTserver es el corazón de cualquier solución Targit. El servidor garantiza la disponibilidad de las fuentes de la compañía en diversas fuentes Targit BI Suite.

TARGIT ANTserver está disponible en dos versiones:

Standard

La tarea principal de la ANTserver es gestionar la comunicación de datos entre el almacén de datos y los clientes Targit. Targit ANTserver es el enlace de conexión para todos los usuarios del sistema. Los servicios a muchos usuarios al mismo tiempo.

Enterprise

El TARGIT ANTserver edición Enterprise es la oferta de esta plataforma para empresas con necesidades complejas y avanzadas. Esta solución permite poner en común todas las bases de datos existentes y de negocios en la misma herramienta Business Intelligence.

Target BI Acceleration.

BI Accelerator ofrece una manera rápida y fácil de crear un diseño de tablas en la herramienta. BI Accelerator combina las sinergias de Targit Suite BI y plataformas de Microsoft SQL Server con habilidades de Data Warehousing.

Pentaho BI

Pentaho es una solución BI completa de open source, que incluye capacidades de consulta, reporting, análisis interactivo, dashboards, proceso ETL y datamining, englobadas en una única plataforma.

Permite la descarga de solución BI completa o sólo de las herramientas especificas que el usuario necesite para su negocio.

Pentaho BI Enterprise Edition incluye:

Pentaho Reporting.

Permite acceder a la información de manera sencilla y distribuirla entre empleados, clientes, proveedores, etc.

Permite además dar salida a la información en diferentes formatos como pdf, html, Excel, formato enriquecido de texto o texto sin formato.

Pentaho Analysis.

Proporciona un análisis rápido e intuitivo de la información que permite a los usuarios no técnicos manejarse por ella con facilidad.

Pentaho Dashboards.

Que permite un control y visibilidad inmediata de los KPI del negocio.

Pentaho Data Integration.

Limpieza e integración de los datos, recopilando éstos del lugar dónde se encuentren y presentando una sola versión de los mismos.

Pentaho Data Mining.

Descubre patrones e indicadores ocultos en la información

En la tabla 3 se comparan las herramientas BI expuestas en cuanto éstas son o no completas, si permiten adquirir sólo parte de la funcionalidad que ofrecen, si incluyen o no proceso ETL en su funcionamiento, si trabajan o no con cubos OLAP y si son o no herramientas open source.

	Solución completa	Dividida en componentes	ETL	OLAP	Open Source
Microsoft Business Intelligence	X		X	X	
QlikView	X		X		
SAP	X		X	X	
ORACLE BUSINESS INTELLIGENCE	X	X	X	X	
TARGIT business intelligence	X	X	X	X	
pentaho open source business intelligence	X	X	X	X	X

Tabla 3. Comparación herramientas BI

2.6 Gestión de proyectos BI

Con demasiada frecuencia los proyectos BI no cumplen con las expectativas puestas en ellos [nase-it.com]. Además suelen darse excesos de coste/tiempo sobre lo previsto. Estos hechos son evitables si se lleva a cabo una correcta gestión del proyecto, para ello debe realizarse una planificación del mismo y contar con una metodología adecuada que ayude en el desarrollo del proyecto y evite, en la medida de lo posible, algunos de los fallos que surgen. Los más habituales de estos fallos son:

- No alinear el proyecto dentro de la estrategia de negocio de la organización
- No realizar al inicio un estudio de viabilidad del proyecto.
- No se tiene en cuenta en tiempo de diseño el alcance en utilización de la herramienta.
- No subcontratar este tipo de proyectos, realizándolos personas de la propia organización sin experiencia suficiente.
- Presupuesto o tiempo dedicado al proyecto escaso.
- Los datos de la organización y que utiliza el sistema BI no están limpios, lo que implicará un proceso ETL más costoso.

La gestión del proyecto consta de las siguientes fases, mostradas en la imagen 18.

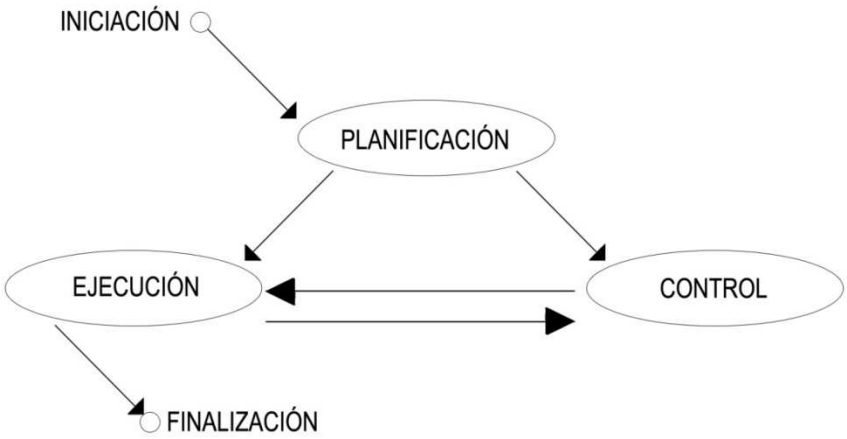


Ilustración 18. Fases de proyecto

Un proyecto BI es un proyecto al fin y al cabo, y por ello, como puede observarse, su gestión se divide en las mismas fases que un proyecto software, aunque como se verá en el presente apartado, estas fases tendrán algunas actividades específicas de un proyecto BI.

Detallando estas fases:

En la **iniciación** del proyecto se define la razón de existencia del mismo, por qué se realiza, a qué necesidad real responde.

En la fase de **planificación** es donde se realiza el grueso del trabajo de gestión del proyecto. En esta fase se preparan los elementos necesarios para hacer posible el proyecto, se adjudican recursos, se dirige y organiza el proyecto para que cumpla con unos plazos en coste y tiempo.

Los objetivos de la planificación son minimizar la duración, el coste y el uso de recursos del proyecto, asegurando plazos y calidad.

Cuando se comienza a hacer la planificación del proyecto hay una serie de aspectos que el jefe de proyecto debe conocer. Éstos son:

- Los objetivos del proyecto.

  Es importante definirlos previo comienzo al proyecto porque cuantificándolos se conocerá más fácilmente si el proyecto ha cumplido o no con todos.

  Una cosa muy importante en los proyectos BI es que los objetivos estén alineados con la estrategia de la empresa, no debe caerse en el error de implantar un producto BI solamente porque estén de moda o porque lo haya implantado la competencia. La necesidad del proyecto ha de estar perfectamente justificada y ha de aportar valor a la organización.

- El alcance.

  El alcance, que normalmente es definido como los procesos que soporta el sistema o las prestaciones que ofrecerá, en el caso de los proyectos BI, el alcance hace referencia a los modelos de negocio que queramos abarcar y los conjuntos de datos necesarios.

También, como en toda definición de alcance, se deberá especificar qué funcionalidad tendrá el sistema y qué queda fuera y dentro del mismo.

- Los riesgos.

El conocimiento y gestión de riesgos es vital es un proyecto BI, por dos motivos. El primero es el coste que su realización implica, que es lo suficientemente elevado como para deber contemplar los elementos que pudieran hacer que el proyecto no se completase con éxito. Y el segundo motivo es que los proyectos BI no se encuentran muy extendidos aún en las organizaciones, y por lo tanto el hecho de decidir realizarlo y fracasar sería predisponer a la organización en la idea de no volver a intentarlo.

El riesgo es un concepto que ha sido tratado por varios autores. Sitkin en 1992 definía el riesgo como “la incertidumbre de tomar una decisión y el resultado que tendrá esa decisión”.

Por su parte el Rowe en 1977 definió incertidumbre como “la ausencia de información concerniente a la situación de tomar una decisión y la necesidad de emitir un juicio para determinar o evaluar la situación, soluciones alternativas o posibles resultados.”

Se define riesgo, más comúnmente, como un evento que, en caso de ocurrir, origina consecuencias que impactan en el proyecto. El riesgo en los proyectos hace que aparezca incertidumbre acerca del nivel de consecución de los objetivos planteados.

Para la toma de una decisión, se busca, se analiza y se procesa la información relevante. Esta información es usada por la persona que tomará la decisión para reducir, eliminar o resolver los riesgos asociados a su decisión. Aceptando que la información constituye una herramienta de reducción del riesgo, se puede concluir que a mejor información, mejor gestión de los riesgos.

Se debe asumir que el riesgo es inherente a la toma de decisiones y que puede provenir tanto de fuentes internas a la organización como externas. Es esencial reconocer la imposibilidad por parte de la organización de eliminar o incluso de aminorar todos los riesgos, especialmente los externos a la misma sobre los que no se tiene poder de decisión.

Sin embargo es necesario llevar una gestión de los riesgos en el proyecto, para posicionar la organización, y desarrollar estrategias para gestionar el impacto de los riesgos y sus consecuencias en términos de objetivos estratégicos de la organización.

Deben elaborarse procesos para identificar, analizar y responder a los riesgos tratando así de minimizar las consecuencias negativas que pudiera ocasionar.

Se puede cuantificar de una manera estructurada los efectos de posibles fallos y priorizar acciones dedicadas a minimizarlos o eliminarlos mediante un Análisis de Modos posibles de Fallo y sus Efectos (AMFE), técnica con la que identifican los riesgos que pudieran surgir.

Detectados éstos, se deben planificar las respuestas, que pueden ser:

- Evitar el riesgo.
- Controlarlo.
- Transferir el riesgo.
- Aceptarlo.

Estas respuestas se incluirán en un plan de contingencia. Se trata de un plan predefinido de reacción ante los riesgos identificados.

Cuando se realiza la gestión de riesgos debe determinarse qué nivel de riesgo es aceptable asumiendo que es imposible afrontar el 100% de los riesgos o hacer que desaparezcan.

Por último, debe tenerse en cuenta que la gestión de riesgos debe revisarse periódicamente.

- Nivel de calidad.

Debe conocerse el nivel de calidad que se exigirá al proyecto teniendo en cuenta las limitaciones, que siempre existen en los proyectos.

Conociendo el alcance, los riesgos, los objetivos y el tiempo y recursos disponibles, se debe fijar un nivel de calidad admisible para el proyecto.

Conocidos estos detalles comienza la planificación en la que se especificará el ámbito del proyecto, el plazo, el coste y los recursos que serán necesarios emplear.

En la planificación se deben especificar:

- Se deben fijar los objetivos, la serie de tareas a realizar.

- La planificación en sí, esto es la secuencia o el orden en que se realizarán las tareas, la dependencia que guardan entre ellas. Las restricciones que pueden darse entre tareas son:

    Restricción final-comienzo: El final de una tarea depende del comienzo de otra.

    Restricción comienzo-comienzo: El comienzo de una tarea depende del comienzo de otra.

    Restricción final-final: El final de una tarea depende del final de otra.

    Restricción comienzo-final: El comienzo de una tarea depende de la finalización de otra.

  La planificación puede representarse con un diagrama PERT o mediante un diagrama PDM, ambos son un método para analizar las tareas involucradas en completar el proyecto, y para determinar el tiempo en completar cada tarea, y así establecer el tiempo mínimo necesario para completar el proyecto total.

- La asignación de recursos donde se asignan recursos a tareas y se define qué personas serán las encargadas de qué actividades. La asignación de recursos puede representarse en una matriz de responsabilidades para que resulte más visual.

- Determinar los puntos de control o hitos en los que se evaluará la marcha del proyecto y se determinará si cumple con lo previsto.

Es importante la planificación del proyecto porque disminuye la incertidumbre y las posibilidades de fracaso.

Después de conocer todos los detalles de la planificación, éstos deben plasmarse en un diagrama de Gantt donde se representarán las diferentes fases, tareas y

actividades programadas y se mostrará la línea de tiempo o de ejecución de las mismas. Se dibuja el diagrama Gantt engloba la planificación completa, ya que PERT no proporciona un único programa de ejecución del proyecto.

La fase de **control** significa que debe revisarse la planificación del proyecto de manera continuada e irse adaptando a los cambios que surjan durante el desarrollo del mismo. La planificación no consiste solamente en un estudio inicial, sino que se debe revisar y modificar según sea necesario.

La **ejecución** trata de la elaboración del proyecto en sí. Puede verse que de la ejecución se vuelve a la fase de control ya que a medida que se avanza en la realización se debe ir verificando si cumple con los resultados esperados en ese punto, de no ser así, de control se pasará de nuevo a la fase de planificación.

La fase de **finalización** es donde se realiza la recogida de datos finales, el último control sobre el proyecto y se determina que se ha cumplido con los objetivos, estudiando qué previsiones no han sido acertadas, en qué puntos de control el proyecto con el calendario, el presupuesto, relaciones… y aprender de estos fallos para próximos proyectos BI.

Debe tenerse en cuenta al evaluar el éxito del proyecto, que éste no debe medirse en tanto cumpla en mayor o menor medida con el tiempo o coste establecido. El proyecto se considerará exitoso si el sistema BI mejora la eficacia en la generación de reportes, en la planificación, en la toma de decisiones y en definitiva aporta valor a la empresa.

Y por supuesto, en la fase de finalización se realiza la entrega del producto BI al cliente.

El esquema de las tareas propuestas para un proyecto BI son las expuestas a continuación. Se dice que son tareas propuestas porque la pauta de tareas que aquí se propone debe ser adaptada en último caso a las necesidades que tenga cada proyecto BI.

1. Planificación del proyecto:
    - 1.1. Definir el proyecto.
    - 1.2. Definir la planificación y la gestión del proyecto.
    - 1.3. Establecer la finalización del proyecto.
2. Arquitectura:

2.1. Revisar los requerimientos de negocio (usuarios, tiempos).
2.2. Definir la arquitectura tecnológica (*hardware*).
2.3. Definir las recomendaciones de configuración.
2.4. Estimar requerimientos de escalabilidad.
2.5. Implementar el hardware y el software.

3. Diseño:
3.1. Desarrollar los modelos de datos.
3.2. Analizar las fuentes de datos.
3.3. Diseñar la base de datos.
3.4. Diseñar el análisis de los usuarios finales.

4. Construcción:
4.1. Revisar el alcance y la planificación.
4.2. Implementar la base de datos.
4.3. Diseñar y desarrollar la integración de datos.
4.4. Cargar y validar la base de datos.
4.5. Construir el análisis de los usuarios finales.
4.6. Probar el sistema.
4.7. Ajustar el rendimiento.

5. Despliegue:
5.1. Entregar la documentación del proyecto.
5.2. Formar a los usuarios.
5.3. Entregar la aplicación.
5.4. Mantener el datawarehouse.

6. Operación:
6.1. Definir los procedimientos de soporte.
6.2. Monitorizar el rendimiento.
6.3. Mantener y mejorar la aplicación

Del esquema propuesto aquí expuesto se realizan dos observaciones:

Arquitectura:

Se debe considerar un adecuado dimensionamiento de la arquitectura a utilizar, si se desatiende este aspecto pueden no aprovecharse óptimamente los recursos y plataformas tecnológicas.

Diseño:

Dentro de la fase de diseño se debe presentar especial atención al modelado dimensional, que ha sido creado especialmente para este tipo de soluciones, y cuyo objetivo es elevar el rendimiento de las consultas.

Un buen modelado puede conducir al éxito del proyecto, del mismo modo que un modelado poco pensado puede conducir al fracaso, ya que si los usuarios no ven una buena performance al analizar la información pueden tender a dejar de utilizar la herramienta.

Existen otras propuestas de tareas a realizar en un proyecto BI [Wise, 2007] que hablan de cinco pasos a seguir para evitar los fallos comentados al principio de este punto y ayudar a garantizar el éxito de los proyectos BI. Estos pasos consisten en la identificación del problema, el conocimiento del uso que se le dará a la herramienta, entender los datos que entrarán en juego, llevar a cabo iniciativas de capacitación y seleccionar una solución:

1. Identificar el problema de negocios.

   Debe identificarse cuáles son las necesidades, fijándose en cuáles son las cosas que no funcionan bien en la organización, y basar la existencia del proyecto en estas necesidades. El proyecto no será bien aceptado por los usuarios finales si éstos no son conscientes de que responde a una necesidad real. (Y es muy importante que los usuarios lo acepten, porque de lo contrario la herramienta resultará inútil).

2. Determinar las expectativas de uso.

   Es necesario estimar en qué medida será utilizada la herramienta, porque debe implementarse conforme a ello. Es un error muy frecuente no prestar atención a este paso y lo que suele suceder después es que la herramienta será utilizada por más personas que las pensadas, lo que causará que el sistema no responda correctamente al elevado número de consultas generadas, lo que provocará, erróneamente, que el usuario pierda fe en el sistema.

3. Comprender la entrega de datos.

   Algunas herramientas incluyen herramientas de mejora e integración de los datos (las encargadas de realizar el proceso ETL visto en el apartado 2.2 de Arquitectura), otras valoran sin embargo que son las propias organizaciones las responsables de sus datos y de la calidad de éstos.

   Tanto en un caso como en otro, es importante comprender la naturaleza de los datos que serán entrada a la herramienta BI, seleccionar qué datos se quieren analizar y realizar una limpieza de éstos.

   Si no se lleva a cabo este paso, el resultado arrojado por la herramienta BI no le será de utilidad a la organización.

4. Implementar iniciativas de capacitación.

   Son muy importantes las iniciativas de capacitación para el éxito del proyecto, ya que contribuye en gran medida a que los usuarios finales acepten el proyecto, y con él los cambios que originará.

   Las iniciativas de capacitación evitarán en gran medida el síndrome de resistencia al cambio, este problema debe ser abordado, porque podría ser razón suficiente para tumbar el proyecto. Una de las formas de abordarlo es mediante las técnicas de capacitación en las que los usuarios entienden y se entusiasman con el proyecto. Estos programas deberían realizarse cercanos a la fecha de implantación, porque si se realizan mucho antes el interés inicial puede desaparecer.

5. Seleccionar una solución vertical u horizontal.

   En este último paso la organización se debe decidir si sería de más utilidad una solución vertical o una horizontal.

   La primera va destinada a organizaciones que necesiten una solución específica para su industria o para un departamento concreto de su organización. Se trata de soluciones más específicas o precisas que las horizontales, que pueden utilizarse en diferentes departamentos o tipos de industria.

La ventaja de las soluciones verticales es que pueden aprovecharse soluciones que fabrican los proveedores de BI para los diferentes tipos de industria, adaptándose a requisitos más concretos que pueden tener unas respecto de otras. Sobre las soluciones horizontales habría que invertir más trabajo para lograr el mismo fin, pero por otro lado, al no basarse en modelos de datos específicos, son más versátiles a las demandas cambiantes de la organización.

Por tanto, conforme a sus necesidades y expectativas de cambio, la organización elige en este último paso, que tipo de solución es la que más le interesa.

Es necesario destacar la importancia de la experiencia en este tipo de proyectos. Se hace necesaria la formación de profesionales dentro de la empresa o bien la determinación de la subcontratación del proyecto, teniendo en cuenta que la formación en este tipo de proyectos es costosa, en el caso de QlikView un curso de formación sobre el manejo de la herramienta supondría a la organización 500 euros por persona y día. Pero es necesario valorar que más costoso resultaría un proyecto mal realizado.

A continuación y para concluir este punto, se van a comentar una serie de factores necesarios para el éxito de proyectos BI. No todos son factores meramente tecnológicos, esto es porque el correcto desarrollo de estos sistemas depende no sólo de la infraestructura técnica, también de la no técnica.

BI es el resultado de relacionar diferentes fuentes de información y mostrarla de manera eficiente para dar soporte a decisiones. Algo primordial en los proyectos BI es la información. Para que un proyecto tenga éxito es necesaria la disponibilidad de la información a todos los niveles de la organización, desde la alta dirección hasta los trabajadores.

Independientemente del tipo de organización y reglas que ésta tenga, hay una serie de factores fundamentales que deben de darse en la organización para que ésta pueda realizar proyectos BI exitosamente:

- Estrategia corporativa cohesiva.

  Todos los procesos de la organización deben estar en línea con la estrategia corporativa de la compañía. De ahí la importancia de que se comunique la estrategia a todo el organigrama de la organización independientemente de

que cada departamento tenga su rol y tenga procesos y objetivos propios encomendados. Todos deben conocer los objetivos comunes. Además los objetivos comunes de un departamento pueden expandirse a diferentes niveles de la organización, posibilitando una respuesta más rápida a posibles cambios del entorno y a las necesidades de los clientes (sean éstos externos o internos a la propia organización).

- Contar con una buena arquitectura tecnológica.

  Un problema derivado de no comunicar la estrategia corporativa a través de todos los niveles de la organización, en este caso de no comunicarla a los desarrolladores de TI, es que los sistemas tecnológicos no sirvan a los intereses de la empresa.

  El no divulgar la información, puede dar lugar a la existencia de múltiples bases de datos con información contradictoria, lo cual podría provocar situaciones erróneas como por ejemplo que diferentes departamentos lleguen a conclusiones diferentes sobre un determinado asunto y sigan caminos diferentes.

  Si la información es compartida por varios departamentos sin un elemento que la vincule, los silos de información no pueden ser resueltos.

  Lo ideal sería que toda la organización compartiera un único sistema de información, pero es posible y habitual que organizaciones de gran volumen necesiten un mayor número de sistemas de información que cubran sus necesidades. Para evitar problemas entre estos sistemas debe desarrollarse una estrategia de desarrollo de sistemas de información y establecerse unos estándares que los controlen.

  BI solucionaría el problema de la información contradictoria, pero para no llegar a seguir caminos diferentes todos los trabajadores deben de conocer los objetivos de la organización.

- Integración del proyecto BI con los procesos de la organización.

  Los procesos de la organización deben estar definidos y debidamente documentados. Debe crearse un mapa de procesos que contenga todos los procesos de la organización, es importante tener definidos correctamente los

procesos para que éstos puedan ser entendidos correctamente y pueda lograrse la consecución de los objetivos comunes de la organización.

Debe tenerse en cuenta que en última instancia la dirección de la organización será la responsable de ofrecer un entorno favorable para la generación de proyectos BI y para el fomento de su posterior utilización.

2.7 El uso de BI en internet

El uso de BI en internet es una demanda en aumento debido a la necesidad, por parte de las grandes empresas, de disponer de sistemas de decisión distribuidos y descentralizados. El uso de las herramientas BI en internet permite analizar datos y adquirir conocimientos necesarios que permiten gestionar los datos localmente, permitiendo a la empresa descentralizar algunas decisiones.

El datawarehouse debe prepararse para permitir su acceso en remoto. Con este fin, las datawarehouse estarán dotadas de interfaces de usuario diseñadas como si fueran navegadores HTML, facilitando el acceso a través de internet.

La información será compartida vía Intranet o vía Extranet, ambos se tratan de canales de comunicación extendidos, que permiten el acceso a usuarios remotos.

Las Intranets son redes internas de las organizaciones, soportadas sobre protocolos TCP/IP, que restringen el uso o acceso a determinadas personas dentro de la organización.

En cambio las Extranets son concebidas para compartir información con otras personas ajenas a la organización, como por ejemplo clientes, socios o proveedores, para conseguir objetivos comunes.

Desde el punto de vista de acceso a la información, las Intranet aportan acceso restringido o privado a la información por parte de los miembros de la organización, mientras que en las extranets el acceso se considera semiprivado, al compartir la información con personas ajenas. Dependiendo de cuál sea el uso que la organización quiera darle a la herramienta BI en internet, elegirá un tipo de acceso u otro a la misma.

Un concepto ligado a las herramientas BI en internet es la minería web.

Web Mining.

Se entiende como Web mining el proceso global de descubrir información o conocimiento potencialmente útil y previamente desconocido a partir de datos en internet.

No es sencillo descubrir información en internet porque no se presenta bajo ningún estándar, no está estructurada y porque su volumen crece rápidamente. Web mining utiliza técnicas de datamining para hallar información relevante de entre toda la disponible en internet. Este proceso puede descomponerse en las siguientes tareas:

- Localización de documentos relevantes en la red. Ésta es la función de los buscadores que extraen contenido basándose en palabras, y de los índices temáticos que clasifican los documentos.
- Extracción de la información del documento independientemente del formato en el que éste se encuentre, HTML, XML, PDF, etc.
- Descubrimiento de patrones generales a partir de sitios web individuales.
- Análisis, validación e interpretación de los patrones.

Dependiendo del objetivo que se persiga, se habla de dos usos de Web mining:

Web mining utilizado para recuperación de información, cuyo objetivo es hallar documentos relevantes. Este tipo de minería web se dispara por consulta y utiliza herramientas estadísticas clásicas.

Web mining como extracción de información, cuyo objetivo es descubrir información relevante en los documentos encontrados. Utiliza herramientas de aprendizaje automático.

Existen varios tipos de Web mining:

- Minería de contenidos.

  Que consiste en la búsqueda de información en diversos tipos de datos en la red, como texto, imágenes, audio, vídeo, metadatos e hiperenlaces. Dentro de este tipo de Web mining se encuadran técnicas como Text mining, Hypertext mining o Markup mining. La utilización de una u otra técnica depende del tipo del documento sobre el que se realiza la búsqueda.

- Minería de estructura.

  Cuyo objetivo es descubrir el modelo subyacente de las estructuras de los enlaces. Consiste en analizar la estructura de los enlaces entre documentos

basándose en la idea de que una página se encuentra reforzada por la cantidad de referencias que posee.

- Minería de uso.

  Este tipo de Web mining se encarga de dar sentido a los datos y de registrar los comportamientos de los usuarios en un sitio web para, por ejemplo, ayudar a la personalización de las aplicaciones

Es necesario el uso de agentes en Web mining. La tarea fundamental de los agentes será la realización de búsquedas proactivas que permitan proporcionar información al usuario o a otros agentes. Para ello los agentes deben ser capaces de realizar la recuperación, el análisis, la manipulación y la fusión de fuentes de información heterogéneas.

La utilización de agentes permite la extracción de información de manera autónoma y la gestión inteligente de datos en Internet y en las Intranets corporativas.

Los agentes pueden clasificarse en:

- Agentes adaptativos. Que pueden adaptarse a los cambios en la red y en la información. Esto es gracias a que son capaces de aprender del usuario y del entorno. Este aprendizaje se base en algoritmos genéticos, redes neuronales, etc.

- Agentes cooperativos/ no cooperativos. Dependiendo de si, en un sistema multiagente, los agentes cooperan o no entre ellos para la consecución del objetivo fijado.

- Agentes racionales. Son agentes cuyo comportamiento se basa en la obtención de beneficios, se comportan de manera utilitarista.

- Agentes móviles. Que son capaces de viajar de forma autónoma por internet.

La principal ventaja del uso de agentes es que éstos pueden realizar solos el trabajo de recolectar información en la red.

La tecnología de los agentes de información corresponde a los campos de la Inteligencia artificial y de las bases de datos.

2.8 Uso y tendencias del Business Intelligence

Referir cuáles han sido las principales ventajas que Business Intelligence ha aportado a las organizaciones es sencillo.

Una de las más importantes es que su uso está acabando con los silos de información existentes entre departamentos, de esta manera toda la información es compartida a lo largo de toda la organización y la gente dispone de la información necesaria para la realización de su trabajo en el momento que lo necesita, tanto es así que la información está dejando de ser enviada a las personas, lo que se conocía como *push*, para ser las personas las que acceden a ella, *pull*, a través de la herramienta BI. Este efecto ha sido documentado como democratización de la información.

La principal utilidad que conlleva este hecho es que a través de este acceso de los usuarios a la información, se les ha facultado para tomar mejores decisiones, lo cual a su vez es el principal reporte de valor para la organización que invierte en esta tecnología.

La tecnología más importante en 2007 fue Business Intelligence [CIO Insight].

Por todo esto el uso de las herramientas BI está en crecimiento [barc, 2007].Muchas organizaciones que empezaron a aplicarlo a nivel departamental y que han comprobado buenos resultados, utilizan ahora BI a nivel corporativo y en las grandes empresas empieza a surgir el concepto de Enterprise Business Intelligence, centros de competencia de Business Intelligence, que son centros dedicados exclusivamente al trato de temas relacionados con el BI y a jornada completa. Algunas de las cuestiones que se tratan en estos centros son:

- Identificación de proyectos potenciales.
- Selección del software más conveniente para proyectos específicos.
- Fijación de políticas de uso del software BI en la organización.
- Coordinación de proyectos de calidad de datos.

De momento solo las grandes organizaciones que llevan muchos proyectos pueden permitirse centros como estos, pero este hecho demuestra nuevamente la importancia de la tecnología BI en las empresas hoy en día.

Una estadística publicada por InformationWeek desvela cuáles son los aspectos más importantes en las herramientas Business Intelligence, ilustrados en la figura 18:

1. Exploración rápida de los datos, consultas, y capacidad analítica.
2. Facilidad de implementación.
3. Facilidad de uso para los trabajadores.
4. Que los usuarios puedan compartir y colaborar con la información.
5. Salida a una amplia gama de formatos.
6. Integración con las aplicaciones de la empresa.
7. Que pueda recoger y analizar datos en tiempo real.
8. Integración con las aplicaciones de escritorio.
9. Permitir una amplia gama de opciones a la hora de visualizar los datos.
10. Poder responder a eventos y proporcionar actualizaciones de datos casi en tiempo real.
11. Poder predecir el comportamiento de clientes, riesgos o resultados de negocio.
12. Amplia oferta de productos ofrecidos.

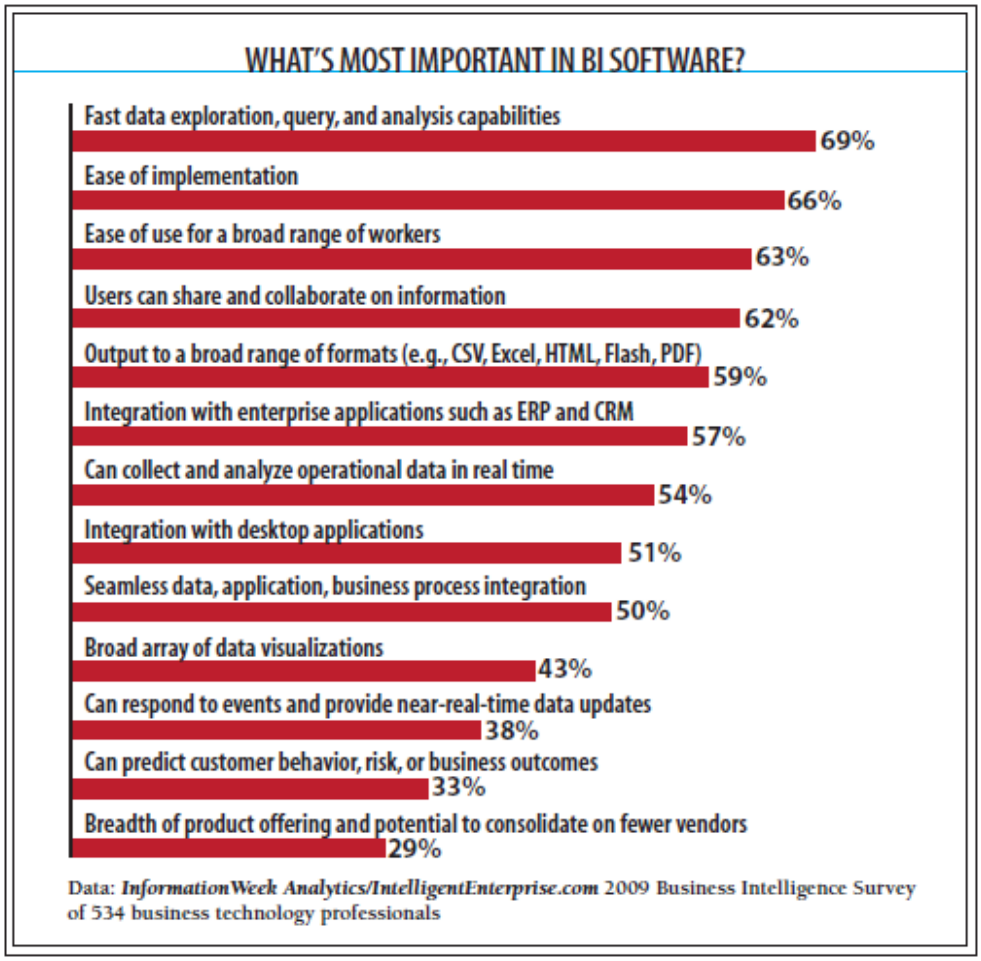


Ilustración 19. Aspectos importantes en BI

Sin embargo, a pesar de las ventajas también ha quedado reflejado en encuestas [barc, 2007] y en diferentes documentaciones que los proyectos BI no se llevan a cabo fácilmente. Según los resultados de The BI Survey, que es una de las grandes encuestas realizadas de manera independiente a los usuarios de Business Intelligence, revela que son los mismos usuarios el principal impedimento para la finalización a tiempo o incluso para la realización misma de los proyectos BI. El nocivo impacto de los usuarios llega a ser más determinante en un fracaso BI que incluso una baja calidad en los datos o un mal software en general.

Este efecto se llama fenómeno de resistencia al cambio, se trata de un comportamiento humano y se da en todos los proyectos informáticos innovadores, como puede ser la implantación de BI en este caso, como pudiera ser también un proyecto de reingeniería, etc. Para combatirlo el jefe de proyecto debe enfrentarse al

problema y nunca a los síntomas que se presentan, teniendo en cuenta que siempre va a tener que afrontar este aspecto, ya que de no encontrarnos en un proyecto como este, un fenómeno de resistencia al cambio es porque realmente los usuarios no son conscientes de la magnitud del proyecto o no tienen predisposición a manejarlo.

Otro problema típico es que las empresas no suelen estar dispuestas a asumir el costo de la implantación de un sistema BI.

Y si realmente no se conoce que tipo de solución se necesita exactamente, las compañías proveedoras de BI pueden venir a confundir más con sus promesas de "podemos hacer cualquier cosa", lo cual suele ser cierto, pero tienden a ignorar las dificultades de la entrega de algunas utilidades.

Y por otro lado, es difícil juzgar los resultados de un proyecto de BI con antelación, porque un buen software produce cambios en la forma de trabajar de las personas, y estos cambios son difíciles de planificar.

Existen una serie de ideas que se piensan serán el futuro del Business Intelligence [InformationWeek, 2009], se trata del análisis predictivo, el análisis en tiempo real y el análisis en memoria:

- El análisis predictivo.

  Esta solución tiene que ver con la minería de datos. Normalmente el datamining ha tenido un enfoque matemático-estadístico. A través de las herramientas BI se pueden realizar proyectos de previsión de demanda, de análisis de riesgos…

  Ya hay empresas, como es el caso de IBM que han comprado soluciones de planificación para integrarlas en sus plataformas BI.

- El análisis en tiempo real.

  El uso por parte herramientas Business Intelligence de los datawarehouse, convierte a éstos en algo más que simples almacenes de información. La información que de ellos se requiere es necesaria para la toma de decisiones, por lo que resulta ineficaz que éstos se carguen una sola vez al día o cada cierto intervalo de horas. Es imprescindible que el dw se encuentre actualizado permanentemente, pudiendo realizar así análisis en tiempo real de los datos, que no conlleven decisiones erróneas.

- El análisis en memoria.

  Actualmente la memoria RAM ya no supone un problema y cada vez (cumpliendo con la ley de Moore) tiene más capacidad. Aprovechando este hecho algunas compañías, entre las que se incluye QlikView, han lanzado soluciones que dejan en evidencia a otras con arquitecturas mucho más complejas y caras. Se almacenan en memoria datos pre-calculados que evitan al usuario tener que esperar por el resultado de la consulta que lanza, si quiera unos segundos.

Algunas publicaciones dirigen a las aplicaciones BI a unirse con las llamadas aplicaciones CPM, de sus siglas en inglés Corporate Performance Management.

Una definición de este concepto dada por Hyperion Solutions es la que sigue:

> *"Un conjunto de tecnologías y procesos que permiten a las Organizaciones traducir estrategias corporativas en planes de acción, monitorear la ejecución de éstos y proveer información que derive en importantes mejoras financieras y operativas"*

El Business Intelligence aplica herramientas analíticas y metodologías a información sumarizada para ser entregada en forma de reportes, tableros de control y escenarios de análisis a los diferentes usuarios de la organización. Las tecnologías CPM funcionan uniendo la información proporcionada por las herramientas BI con las estrategias financieras y operativas de la empresa, lo cual lleva a tomar acciones concretas.

Las tecnologías de BI y CPM obtienen su mayor impacto y beneficio en su convergencia, pero es muy importante mencionar que esta convergencia se debe dar bajo un marco conceptual bien definido y una arquitectura tecnológica integrada.

Los últimos artículos sobre las tendencias del BI señalan que a medida que el mercado de Business Intelligence evoluciona aparecen cambios que darán forma a la próxima generación de BI [Stephen B.]. Se distinguen cuatro tendencias importantes: Cloud Computing y la virtualización, Pervasive BI, la migración del procesamiento de Data Marts externos hacia un enfoque integral de Data Warehouse y la integración de tipos de datos no tradicionales. Exploraremos las implicaciones de estas tendencias en el desarrollo de las soluciones de un Data Warehouse.

3. Aplicación BI

Como parte del PFC se ha realizado una aplicación BI utilizando la plataforma de desarrollo QlikView 9.0 Personal Edition.

En los siguientes puntos se detallarán, los datos de entrada a la aplicación, la herramienta BI comercial escogida y el manual de usuario para su manejo.

3.1 Datos de entrada

Los datos de entrada al documento QlikView, han sido tomados de una base de datos Access, desarrollada al efecto de crear unos datos que sirvieran para el objetivo del presente capítulo, que es mostrar cómo funciona una herramienta BI comercial.

La base almacena datos de patologías, edificaciones, tratamientos, etc. Por lo que será analizada desde el punto de vista de una organización ficticia, dedicada a la arquitectura, y que utilizaría el sistema BI para obtener información acerca del diagnóstico de patologías en estructuras arquitectónicas y de los proyectos al efecto que en la organización se han desarrollado.

El diagrama E/R correspondiente a la base de datos es el que sigue en la imagen 20

Esquema E/R

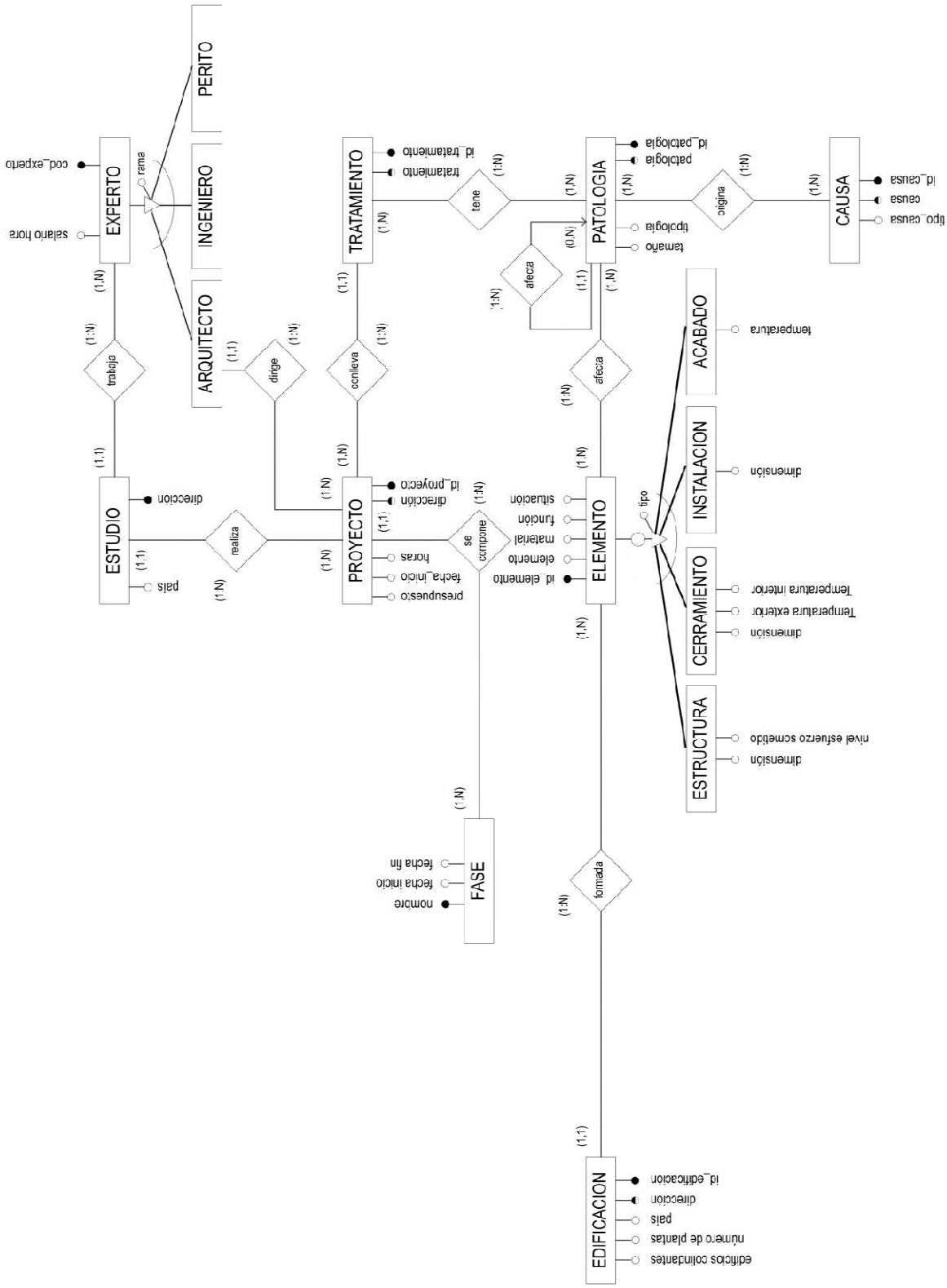


Ilustración 20. Esquema E/R

Las tablas de las que está compuesta y sus respectivos campos son:

1. La tabla Patología, que contiene información relativa a patologías.

PATOLOGÍA	
Id_patología	Identificador unívoco de una patología en la base de datos.
Patología	Nombre por el que se conoce a la patología
Tipología	={física, química, mecánica}. Corresponde al tipo de acción que provoca la patología y de la propia evolución de la patología. Sirve para la clasificación de la misma.
Tamaño	Dimensiones que adquiere la patología, en caso de que proceda.

Tabla 4. Tabla Patología

2. La tabla Causa, que contiene información relativa a las causas que pueden conllevar a una estructura a sufrir una determinada patología.

CAUSA	
Id_causa	Identificador unívoco de una causa en la base de datos.
Causa	Nombre por el que se conoce a la causa u origen de la patología.
Tipo_causa	Corresponde al tipo de acción que provoca la causa en sí misma.

Tabla 5. Tabla Causa

3. La tabla Tratamiento, que contiene información relativa a los tratamientos o acciones que pueden llevarse a cabo para solucionar una determinada patología.

TRATAMIENTO	
Id_tratamiento	Identificador unívoco de un tratamiento en la base de datos.
Tratamiento	Nombre por el que se conoce a la causa u origen del tratamiento.

Tabla 6. Tabla Tratamiento

4. La tabla Edificación, que contiene información de los edificios para los que se ha realizado un proyecto de diagnóstico.

EDIFICACIÓN	
Id_edificación	Identificador unívoco de una edificación, sobre la que existe una patología, en la base de datos.
Dirección	Dirección del edificio
País	País en el que se encuentra el edificio.
Número de plantas	Número de plantas construidas del edificio.
Edificios colindantes	={si, no}. Booleano que índica la existencia o no de edificios colindantes a la edificación afectada.

Tabla 7. Tabla Edificación

5. La tabla Proyecto, contiene información de los proyectos de diagnóstico realizados.

PROYECTO	
Id_proyecto	Identificador unívoco del proyecto en la base de datos.
Dirección	Dirección de la edificación sobre la que se realiza el proyecto.
Fecha_inicio	Fecha de comienzo
Horas	Horas de realización del proyecto.
Presupuesto	Coste total del proyecto.

Tabla 8. Tabla Proyecto

6. La tabla Estudio, contiene información de las diferentes oficinas de la organización.

ESTUDIO	
Dirección	Dirección en la que la organización tiene un estudio.
País	País del mismo

Tabla 9. Tabla Estudio

7. La tabla Elemento, guarda la información de los diferentes componentes de una edificación, sobre los cuales surgen las patologías.

ELEMENTO	
Id_elemento	Identificador unívoco del elemento en la base de datos.
Elemento	Nombre por el que se conoce al elemento.
Material	Material del que está constituido
Función	Función que cumple como parte de la estructura del edificio.
Situación	Ubicación en la edificación.
Tipo	={Estructura, cerramiento, instalación, acabado}. Corresponde al tipo de elemento, sirve para la clasificación de los mismos. Éstos son las tablas hija de la tabla 'elemento' en el diagrama E/R. Para su implementación se ha decidido migrar los atributos a la superclase, e identificarlas con este campo tipo.
Dimensión	Tamaño del elemento, en el caso que corresponda.
Nivel de esfuerzo	Nivel de resistencia o fuerza externa a la que está sometido el elemento, en el caso que corresponda.
Temperatura interior	Temperatura interior de la edificación en la ubicación del elemento, en el caso que corresponda.
Temperatura exterior	Temperatura exterior a edificación en la ubicación del elemento, en el caso que corresponda.
Temperatura	Temperatura soportada por el elemento, en el caso que corresponda.

Tabla 10. Tabla Elemento

8. La tabla Experto, que contiene información de los diferentes profesionales que trabajan en la compañía.

EXPERTO	
Cod_experto	Identificador unívoco del empleado en la base de datos.
Rama	={Arquitecto, Ingeniero, Perito}. Ámbito de dedicación del experto.
Salario hora	Retribución económica por hora trabajada al experto.

Tabla 11. Tabla Experto

9. La tabla Fase, guarda contiene información de las fases de las que está compuesto un proyecto.

FASE	
Nombre	={toma de datos, análisis, ideación de la solución, delineado de la solución}. Nombre por el que se conoce a la fase por la que pasa un proyecto.
Fecha inicio	Fecha de entrada del proyecto en la fase.
Fecha fin	Fecha de conclusión del proyecto en la fase.

Tabla 12. Tabla Fase

Entendiendo la base de datos desde el campo de la arquitectura se dice que:

Una determinada edificación está formada por elementos, éstos pueden ser parte de la estructura, un cerramiento, un elemento de instalación o un elemento de acabado.

Son estos elementos los que se ven afectados por las patologías. Éstas están originadas por una causa externa y siempre tienen al menos un tratamiento que las soluciona.

La organización llevaría a cabo proyectos de diagnóstico, que encontrarán una solución o tratamiento a la patología. Los proyectos pueden dividirse en fases, para su mejor gestión.

Son los diferentes estudios de arquitectura, los que llevan a cabo los proyectos. En estos estudios trabajan una serie de profesionales o expertos, de las ramas de la arquitectura, ingeniería y peritaje, siendo el arquitecto el encargado de dirigir y firmar los proyectos.

3.2 Plataforma BI

La plataforma BI elegida ha sido QlikView. Ésta es la solución de Business Intelligence creada por la empresa QlikTech®.

La versión utilizada es QlikView 9.0 Personal Edition. En esta versión no existen limitaciones para el desarrollo, pero tiene la excepción de que no podrán abrirse documentos QlikView creados por otros usuarios, ya que el documento se guarda con una clave, compuesta por información del usuario y el ordenador en el que está trabajando. Esto significa que, en teoría, no puede abrirse un documento QlikView desde otro ordenador. No obstante, al intentar abrirlo se pregunta al usuario si desea recuperar el archivo. Respondiendo sí, éste puede ser abierto, y se genera una nueva clave de usuario para el documento. El máximo número de cambios de clave permitidos para ésta versión de QlikView es cuatro.

QlikTech® creó y patentó para QlikView, una nueva tecnología de lógica asociativa en memoria, gracias a la cual la aplicación es capaz de efectuar cálculos en tiempo real, permitiendo además navegar por la información de una manera más intuitiva que con otras soluciones.

Las razones que llevan a elegir la plataforma BI QlikView para este PFC son:

- Innovadora arquitectura que rompe con los sistemas de BI tradicionales. Este aspecto se verá más en profundidad.
- Por segundo año consecutivo ocupa el primer puesto en la encuesta BI Survey 9[8], superando incluso a las plataformas Hyperion (Oracle), Business Object (SAP) y Cognos (IBM).

[8] La encuesta BI Survey es la mayor encuesta realiza a usuarios de Business Intelligence. Es realizada cada año por Business Aplication Research Center (BARC). BI Survey 9 se realizó

- La versión QlikView utilizada se trata de una solución completa, que se encuentra además disponible desde su sitio web de manera gratuita, requisito indispensable en la elección de la herramienta para este proyecto, ya que las licencias de este tipo de productos son altamente caras.

- La apuesta de QlikView por una nueva tecnología que soportarse la plataforma BI residiendo la información en memoria ha sido ampliamente aceptada y así queda reflejado en el MQ de Gartner, donde QlikView se posicionaba en 2008 en grupo de los *visionarios*, mientras que en 2010 se posiciona en el grupo de los *rivales* rozando el cuadrante de los *líderes*.

3.2.1 Arquitectura QlikView

La arquitectura de los sistemas BI en disco tradicionales consiste, como se ha visto, en extraer los datos (proceso ETL) y almacenarlos en datawarehouses y en estructuras multidimensionales (los cubos), para que éstos sean accedidos por la herramienta y así pasen al análisis de los usuarios.

QlikView simplifica la arquitectura y se soporta sobre su tecnología asociativa residente en memoria. Soporta funciones ETL para extraer, transformar y cargar los datos de diferentes orígenes de datos directamente a memoria.

Los datos residirán en RAM, prescindiendo de crear cubos OLAP multidimensionales y sin la necesidad de un datawarehouse, porque estas son estructuras costosas, lentas de crear y cuyo mantenimiento conlleva tiempo (como se vio, si cambian los datos, habría que actualizar el datawarehouse y crear de nuevo el cubo).

El tipo de arquitectura residente en memoria propuesto por QlikView es posible gracias a las soluciones hardware que existen hoy en día de 64 bits, con gran capacidad de memoria y procesadores multinúcleo que permiten realizar cálculos paralelos con un gran rendimiento.

Al apoyarse QlikView sobre una arquitectura común, sirve no solo dentro de la organización sino también entre sus redes de clientes, distribuidores, proveedores, etc.

mediante un estudio online sobre un espacio muestral de 2000 usuarios y analizando 23 productos BI.

Por otro lado, como los datos se conservan en memoria, se eliminan las limitaciones de los cubos en disco y se puede acceder más rápidamente a los datos, así por ejemplo la respuesta a cualquier selección que conlleve un cálculo se realiza de manera mucho más rápida.

Aprovechando la velocidad de residir en memoria QlikView permite realizar tareas, que las soluciones tradicionales sobre disco no permiten como:

- Poder modificar el análisis a capricho del usuario, pudiendo realizar cálculos en tiempo real.
- Entender las relaciones asociativas disponiendo del conjunto total de los datos.
- Realizar análisis complejos como análisis de conjuntos, que para realizar las soluciones tradicionales consumirían mucho tiempo.
- Llevar a cabo análisis y presentarlos con rapidez y de forma simultánea a usuarios conectados en diferentes localizaciones sin problemas de carga.

Además al residir en memoria se eliminan problemas que conllevan las soluciones tradicionales de disco y que son:

- Incapacidad de crear análisis con rapidez.
- Acceso lento a los datos.
- Dependencia del departamento de informática de la organización por parte de los usuarios para crear o modificar análisis.

Resumiendo, QlikView se trata de una solución BI completa, soportando las herramientas BI tradicionales en una misma arquitectura. Lo que permite consolidar los datos, realizar búsquedas asociadas, realizar cálculos y presentar visualmente los resultados mediante cuadros de mando, análisis e informes, todo desde una misma versión.

3.2.2 Lógica Asociativa.

La manera que tiene QlikView de operar con sus datos ha sido denominada AQL (Associative Query Logic). Se trata de una tecnología de desarrollo de bases de datos aplicada en QlikView y patentada por QlikTech, que la define como sigue:

> *"Cuando se comprime la información, se comprime y normaliza al máximo para que no haya información redundante. Cada valor único para todos los datos se almacena una sola vez y se referencia a través de punteros. Por ejemplo, si el primer registro de una fuente de datos incluye el campo 'coche rojo' y la segunda incluye el valor 'coche negro', se almacena 'coche' una sola vez. En lugar de almacenar dos veces 'coche', un contador asociado a un puntero referencia el incremento de ese valor."* [QlikTeach, 2006]

Esta tecnología consiste en mantener en RAM una base de datos no relacional, sino asociativa. La lógica asociativa permite a los usuarios manejar los datos de la misma manera que piensan, es decir, de forma asociativa, permitiéndoles ver al instante las conexiones y relaciones entre los datos, por sí mismos.

La principal ventaja de la lógica asociativa es que los datos fuente son retenidos y puestos a disposición inmediatamente. El resultado en QlikView es su potente capacidad analítica que permite explorar y descubrir relaciones no obvias en los datos, aportando mayor conocimiento al usuario sobre éstos.

3.3 Manual de usuario

3.3.1 Interfaz ODBC

El primer punto para comenzar a realizar la aplicación QlikView es realizar la importación de los datos que se quieren visualizar, sobre los que se quiere obtener información. Los datos pueden encontrarse en diferentes formatos y en sitios diferentes y sin embargo ser necesarios todos para poder realizar un análisis en conjunto y poder tomar una decisión.

Lo más común es que los datos se encuentren contenidos dentro de una base de datos, en este caso, para importar los datos a QlikView es necesario tener instalada una interfaz ODBC (Open Data Base Conectivity). QlikView funcionará con drivers ODBC de 32 bits y de 64 bits; en un sistema de 32 bits sólo podrá ejecutarse un driver ODBC de 32 bits, mientras que en un sistema de 64 bits permite manejar ambos drivers, tanto de 32 como de 64 bits.

Algunos drivers son instalados con el sistema operativo, como es el caso del driver MS Access Database, que será el utilizado en esta aplicación para importar la base de datos BDPatologias.accb.

Pueden comprobarse los drivers instalados en el sistema a través del diálogo ODBC de la siguiente manera:

- Abrir el panel de herramientas administrativas (Inicio -> Panel de control -> Herramientas administrativas)

- Abrir el diálogo 'Orígenes de datos ODBC'. En la pestaña ODBC Drivers se encuentra la lista de los drivers instalados. Si en dicha lista no se encuentra el driver necesario, no se podrá importar la base de datos. En este caso se tienen dos opciones: conseguir e instalar el driver o transformar los datos que se desean importar a un archivo legible por QlikView, como podría ser por ejemplo a un fichero access.

Una vez comprobado o conseguido el driver ODBC necesario, se debe declarar la base de datos a importar como fuente de datos ODBC. Al crearla debe decidirse si esta fuente será de usuario o de sistema. La diferencia es que a una fuente de datos declarada de usuario sólo podrá accederse con el id de usuario correspondiente, por lo

que si se quiere que a la fuente de datos tengan acceso diferentes usuarios debe crearse como fuente de datos de sistema. En ambas opciones, para declarar la base de datos como origen de datos ODBC se procede de la siguiente manera:

- Abrir 'Administrador de orígenes de datos ODBC' (Inicio -> Panel de control -> Herramientas administrativas -> Orígenes de datos ODBC)

- Seleccionar la pestaña DSN de usuario, para crear una fuente de datos de usuario, o DSN de sistema, para crear una fuente de datos de sistema.

- Al seleccionar 'Agregar' se abrirá el diálogo mostrado, en la figura 21, 'Añadir fuente de datos', en el que se encuentra la lista de todos los drivers instalados en el sistema. Se selecciona el driver correcto y se pulsa 'Aceptar'

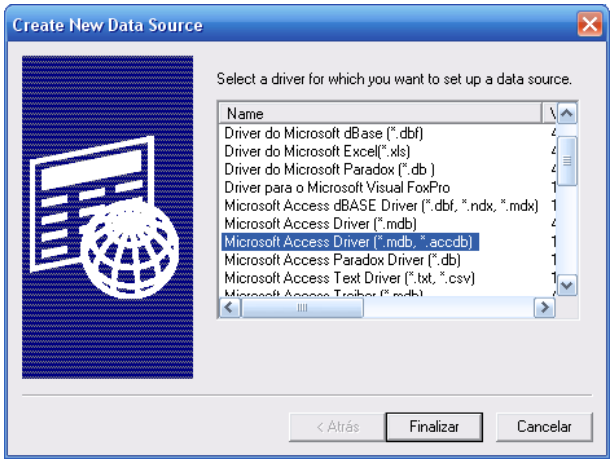


Ilustración 21. Diálogo Añadir fuente de datos

- Se abre entonces el diálogo correspondiente al driver seleccionado, como ilustra la imagen 22.

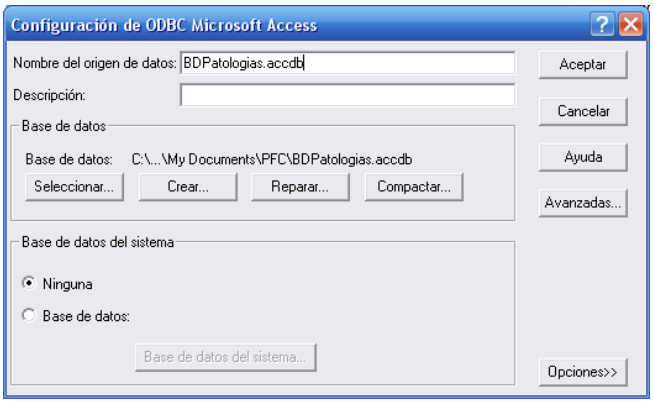


Ilustración 22. Diálogo configuración de ODBC

En este diálogo deberán especificarse:

- El nombre del origen de datos: Se dará el mismo nombre que tenga la base de datos.

- La ruta de la base de datos, pulsando en botón 'Seleccionar'

3.3.2 Importar datos

El diálogo 'Editor de script' (Desde el documento QlikView: Archivo -> Editor de script), representado en la figura 23, permite introducir y ejecutar un script que conecte la aplicación con una fuente de datos ODBC o con ficheros planos de diferentes tipos, para importar la información necesaria.

Puede generarse el script manualmente o a través de la funcionalidad aportada por QlikView.

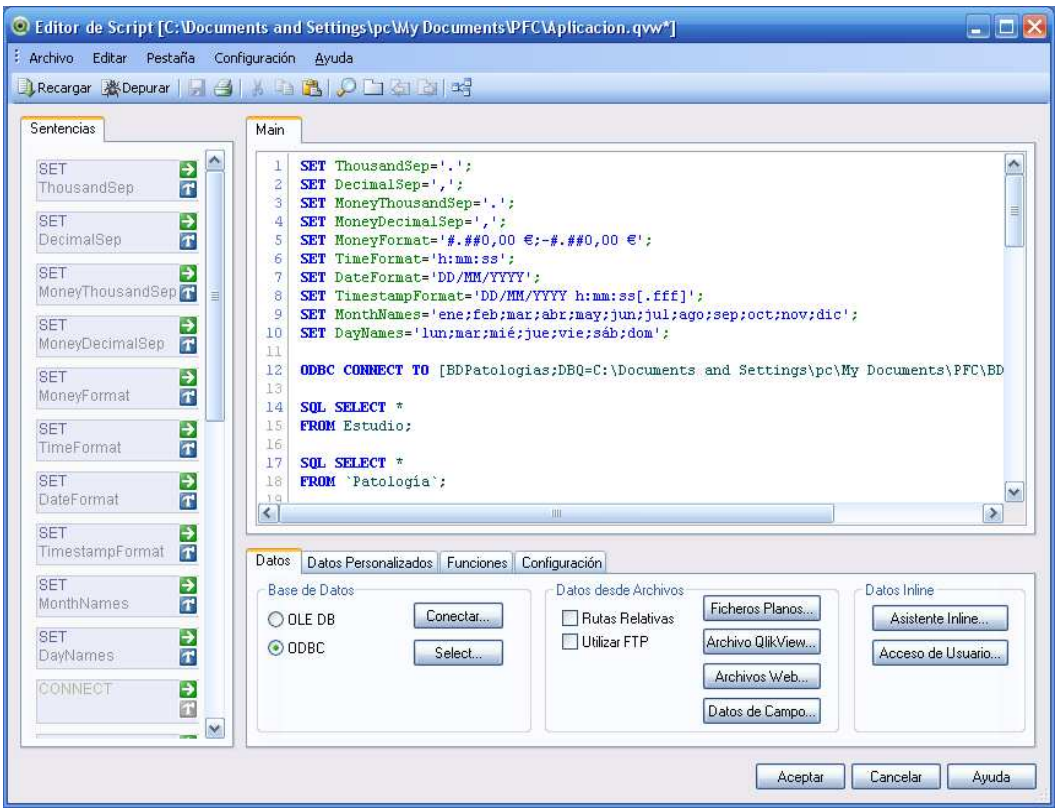


Ilustración 23. Editor de script

El editor de script está divido en diferentes sub-paneles:

- El panel *Main* contiene el script mismo, este panel puede dividirse en varias pestañas para el mejor manejo del código por parte del usuario.

- El *panel de sentencias* que aparece a la izquierda es un esquema de las sentencias contenidas en el script, que facilita la movilidad por el mismo.

- El panel inferior que contiene las pestañas *Datos, Datos personalizados, Funciones y Configuración:*

    - La **Página de datos** permite seleccionar los parámetros de la importación de los datos a QlikView.

      <u>Base de datos</u>

      Mediante estos parámetros se establece una conexión con una fuente de datos, a través de una interfaz ODBC u OLE DB.

      Como se contó en el apartado anterior, es necesario disponer del driver ODBC u OLE DB que soporte la base de datos antes de poder recuperar los datos.

      Para importar los datos de la fuente de datos se selecciona que tipo de conexión a ésta se quiere: ODBC/OLE DB y se pulsa conectar. En este momento se abre el diálogo *Crear sentencia select*, donde se definen las tablas y campos de tablas que se cargarán de la fuente de datos. Este panel se explica en el apartado 'Panel crear sentencia select' del presente documento.

      <u>Datos desde archivos</u>

      Es posible también importar datos en QlikView que no provengan de una fuente de datos definida como tal. Pueden seleccionarse ficheros planos o incluso archivos qlikview (.qvw). Estos archivos pueden ser seleccionados desde un servidor ftp, seleccionando un servidor de la lista y pulsando *Aceptar*.

      Cuando se importan datos desde archivos puede ser necesario editar la configuración de esos datos para indicar a QlikView cómo debe

interpretar los datos de entradas, esto se hace a través de la *Página Tipo*, que se detallará más adelante.

- La **Página de Datos personalizados** permite introducir a QlikView datos procedentes de otras fuentes personalizadas. Existe la posibilidad de programar interfaces personalizadas para tipos de fuentes de datos, no cubiertos por las interfaces de archivo tradicional ODBC/OLE DB, el caso más típico es importar datos disponibles vía Web Services.

  Connect... Abre el diálogo para conectar con la fuente de datos personalizados.

  Seleccionar... Abre el diálogo donde seleccionarse los archivos de la fuente de datos seleccionada.

- La **Página Funciones** contiene herramientas para editar y moverse por las funciones del script.

  Permite editar funciones mediante las opciones:

  - Categoría de la función. Donde se selecciona el tipo de función de la lista.
  - Nombre de Función. Donde se selecciona una función de la categoría seleccionada.
  - Pegar. Pega la función en el script.

  Posteriormente habrá que editar la función pegada para que se adapte totalmente a la funcionalidad que se desea, pero mediante esta herramienta se agiliza el trabajo de editar las funciones del script.

- La **Página Configuración** permite editar medidas básicas de seguridad para el script de QlikView. Pueden configurarse aspectos como:

  - Abrir la base de datos en modo lectura/escritura, permitiendo así modificaciones sobre la base de datos.
  - Permitir ejecutar programas externos.

- Codificar el userID y su correspondiente contraseña al realizar una conexión a la base de datos.

3.3.2.1 Panel Crear Sentencia Select

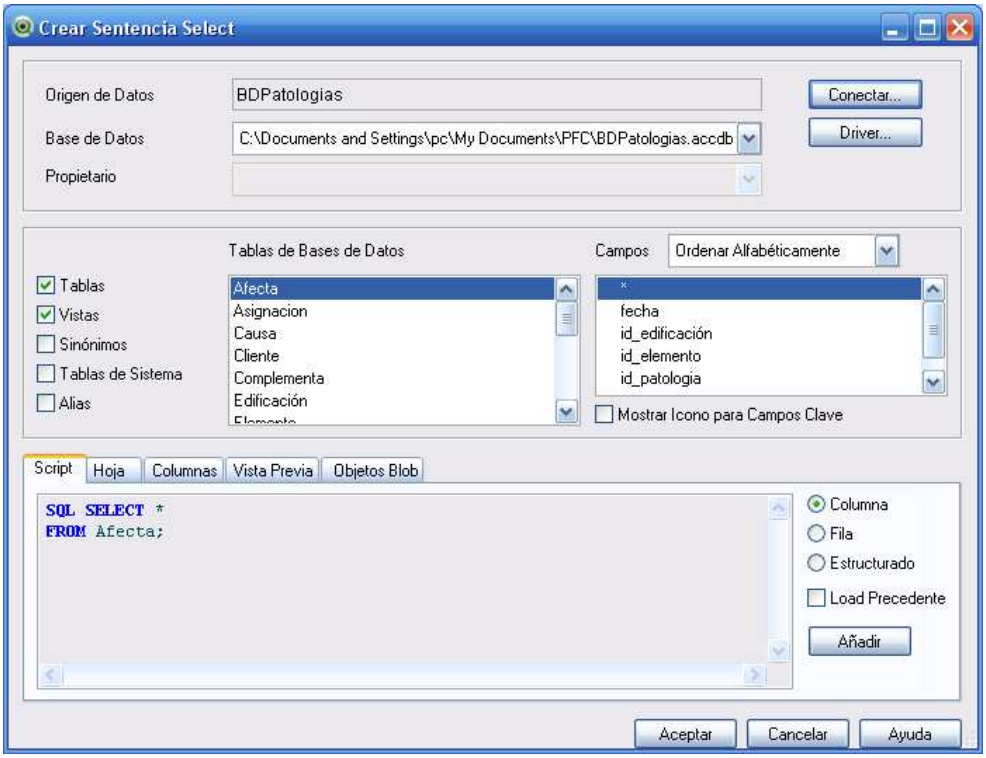


Ilustración 24. Diálogo Crear Sentencia Select

Este dialogo, mostrado en la figura 24, consta de tres paneles.

El panel superior contiene información acerca de la base de datos.

Origen de datos. Aparece el nombre del ODBC/OLE DB seleccionado.

Base de Datos. Donde se muestra la lista de bases de datos disponibles dentro de la fuente de datos definida.

Conectar… Pulsando conectar se accede al diálogo propiedades de vínculo de datos donde puede elegirse una fuente de datos definida o crear una nueva conexión a una fuente de datos.

El panel de en medio sirve para seleccionar tablas y campos de la base de datos.

Tablas de base de datos. Se trata de una lista de todas las tablas contenidas en la base de datos. Puede seleccionarse qué tipo de tablas se quiere que aparezcan en la base de datos, activando o desactivando las casillas de la izquierda, las más comunes son:

- Tablas. Si se selecciona, aparecerán en la lista las tablas regulares de la base de datos.
- Vistas. Si se selecciona, aparecerán en la lista las vistas de la base de datos.
- Tablas del sistema. Si se selecciona, se añadirán a la lista las tablas del sistema.

Campos. Pulsando sobre el nombre de una tabla se cargan en este panel, los atributos de los que está compuesta la tabla. Pueden importarse tablas completas o sólo algunos de sus campos, para hacerlo se selecciona ‘ * ‘, si quieren importarse todos, o el conjunto de campos a importar y se pulsa *Añadir* en panel inferior.

En el panel inferior se muestran las sentencias select que se han ido generando en las selecciones de inserción de datos realizadas. Al pulsar Añadir se ejecutará la sentencia select generada hasta el momento y comenzará la edición de una nueva sentencia select sin abandonar el diálogo.

Puede controlarse el formato de las sentencias select con la selección de los siguientes botones de opción:

- Columna. Si se selecciona esta opción, se generará una sentencia select por cada nombre de campo.
- Fila. Si se selecciona esta opción, se generará la sentencia select en una sola línea.
- Estructurado. Si se selecciona ésta opción, se generará la sentencia select en filas estructuradas.

Una vez añadidos todos los campos/tablas deseados se pulsa *Aceptar* y se vuelve al diálogo Editor de script, donde el panel main aparecen ya las sentencias 'load' generadas.

Botón recargar. A continuación se pulsa sobre recargar, para que los datos asociados a las sentencias representadas en *Main*, quedan añadidos al archivo qvw.

Las sentencias de script generadas para este proyecto han sido:

```
ODBC CONNECT TO [BDPatologias;DBQ=C:\Documents and Settings\pc\My
Documents\PFC\BDPatologias.accdb];
SQL SELECT *
FROM Estudio;
SQL SELECT *
FROM `Patología`;
SQL SELECT *
FROM Tratamiento;
SQL SELECT *
FROM Causa;
SQL SELECT *
FROM Trata;
SQL SELECT *
FROM Afecta;
SQL SELECT *
FROM Origina;
SQL SELECT *
FROM Elemento;
SQL SELECT *
FROM Proyecto;
SQL SELECT *
FROM Experto;
SQL SELECT *
FROM Asignacion;
SQL SELECT *
FROM `Edificación`;
SQL SELECT *
FROM Estudio;
```

La sentencia ODBC CONNECT TO, conecta al documento QlikView con la base de datos.

Las sentencias SQL SELECT, importan los datos de la base.

3.3.2.2 Página Tipo

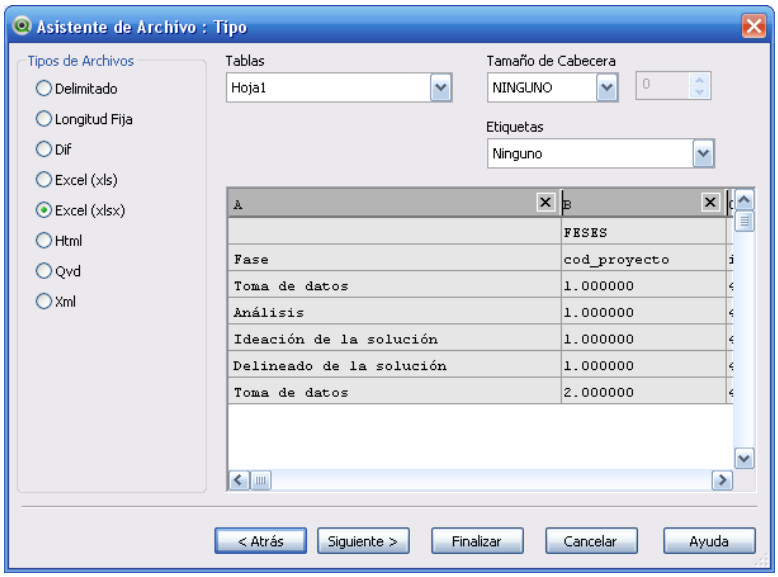


Ilustración 25. Página Tipo

A través del diálogo *Asistente de archivo: tipo*, mostrado en la figura 25 pueden editarse las configuraciones relativas al tipo de archivo o tipo de tabla. Cuando se importa un archivo plano QlikView trata de interpretarlo y se fijan una configuración dependiendo del tipo de archivo que se trate, a través de la página puede cambiarse dicha configuración.

Para ello se comienza seleccionando en los botones 'Tipo de archivo' el grupo al que pertenece el archivo de datos que se va a importar. Dependiendo del tipo de archivo seleccionado, podrán configurarse diferentes elementos, los más comunes son los siguientes:

- Tablas. En caso de importar un archivo con varias tablas se selecciona en este desplegable la tabla a tratar.

- Tamaño de la cabecera. Permite omitir la cabecera del archivo, que corresponde a la primera fila del mismo o al número de bytes indicados.

- Etiquetas. Permite definir si los nombres de los campos, las cabeceras de columna, se encuentran contenidos en el archivo, concretamente en la primera fila de la tabla. Si es así debe seleccionarse la opción *Etiquetas incrustadas*, o si por el contrario, no se tienen los nombres de las tablas, en cuyo caso se seleccionará la opción de etiquetas *Ninguno*.

- Analizar posición fija. Permite definir anchos de columna para archivos que contengan registros de longitud fija.

- Delimitador. Permite especificar los delimitadores (tabuladores, coma, punto y coma, espacios, o la definición de un delimitador propio) y el esquema de entrecomillado a emplear (estándar, MSQ o ninguno) en archivos de tipo delimitado.

3.3.3 Lógica y selecciones

En la hoja2 del archivo .qvw se han especificado algunas nociones sobre cómo navegar por los datos en QlikView. No obstante, en el presente apartado se expondrá con mayor detalle la lógica y selecciones de datos en QlikView.

3.3.3.1 Selección de valores.

Para seleccionar un valor sólo hace falta hacer click sobre él y quedará seleccionado. Pueden seleccionarse varios valores de campo a la vez, para ello basta con arrastrar el cursos por los valores a seleccionar, si estos valores no se encuentran contiguos se selecciona el primer valor y para cada selección adicional se pulsa Ctrl+click.

De la misma manera que se realiza la selección de campos, se realiza también la eliminación de selección de los mismos, haciendo click sobre un campo seleccionado, éste se deseleccionará.

La selección de un campo puede conllevar selecciones automáticas en otros objetos de hoja tales como gráficos, cuadros de lista, tablas, etc. Estos campos relacionados se distinguirán como tal porque se colorean de color blanco, mientras que los que no guardan ninguna relación con la selección actual permanecen en tono gris.

El estado de los campos, dada una selección, se identifica pues por su color. Los códigos de color aplicados en QlikView son los siguientes:

Color	**Estado**
Verde	Seleccionado
Blanco	Opcional
Gris	Excluido
Azul	Bloqueado

Si la selección actual es múltiple, es decir, se tienen varios campos seleccionados y por lo tanto marcados en verde, y se realiza una selección nueva de los valores opcionales, es posible que alguno de los valores de la selección actual cambien su valor de seleccionados a excluidos.

Una selección múltiple es interpretada por defecto como un OR lógico, lo que significa que QlikView buscará una solución relacionada con uno o más de los valores seleccionados. Puede establecerse una selección múltiple de tipo AND lógico si se quiere que QlikView busque soluciones relacionadas con todos los valores de la selección actual. Para ello es necesario configurar el Cuadro de lista en un modo lógico AND (Objeto -> Propiedades -> General -> Marcar 'Modo And'

Bloquear valores de campo seleccionados.

Activada una determinada selección, si se selecciona un valor excluido, la selección actual puede perderse. Si interesa mantener dicha selección, ésta puede bloquearse en el cuadro de lista desde el menú *Objeto* y seleccionando la opción *Bloquear* o mediante el teclado pulsando Ctrl+L. Al bloquearse, la selección actual pasará a visualizarse en color azul.

Se pulsará la opción Desbloquear desde el mismo menú para eliminar la opción de bloqueo o mediante teclado pulsando Ctrl+U.

Debe tenerse en cuenta que cuando se intente seleccionar un valor que no sea compatible con una selección bloqueada, esta selección fallará.

3.3.3.2 Búsqueda de valores

Pueden realizarse selecciones mediante búsquedas de texto en el cuadro de búsqueda, mostrado en imagen 26, que aparece al hacer click en la cabecera del cuadro de lista.

Ilustración 26. Cuadro de búsqueda

Para realizar la búsqueda se introduce en el cuadro de búsqueda la cadena deseada. La cadena de búsqueda no es sensible a mayúsculas. QlikView mostrará todos los valores de campo que cumplan los criterios de búsqueda de la cadena. El cuadro de búsqueda se cierra automáticamente cuando se pulsa ENTER. Puede cancelarse la operación de búsqueda pulsando ESC.

QlikView contempla cuatro tipos de búsqueda: búsqueda de texto, búsqueda difusa, búsqueda numérica y búsqueda avanzada:

Búsqueda de texto

Es la forma de búsqueda más simple. QlikView busca valores de campo que coinciden con la cadena de texto introducida.

A la cadena de texto pueden añadírsele caracteres comodín, caracteres arbitrarios en un lugar de la cadena. Utilizando el símbolo ***** se indica la existencia de cero a N caracteres comodín, y con el símbolo ? se indica la existencia de un único carácter arbitrario. Así por ejemplo opciones de búsqueda serían:

cadena. De manera predeterminada QlikView sitúa la cadena entre asteriscos, lo que indica que se buscan valores de campo que contengan ‘cadena’.

cadena*. Buscará valores de texto que comiencen por ‘cadena’

Búsqueda difusa

Comenzando la búsqueda de texto con el símbolo ~, se indica que quiere realizarse una búsqueda difusa. La búsqueda difusa ordena los valores, mientras se escribe, por grado de semejanza con la cadena de búsqueda, situándose los valores más coincidentes al principio de la lista. Si se pulsa INTRO se selecciona el primer valor de la lista.

Búsqueda numérica

Pueden realizarse búsquedas con valores numéricos comenzando la cadena con un operador relacional (>, >= , < , <=) y escribiendo a continuación el valor numérico deseado.

Ejemplos de búsquedas numéricas podrían ser:

>10<100 Busca valores comprendidos entre 10 y 100.

<10>100 Busca los valores menores de 10 ó mayores de 100.

Búsqueda Avanzada

Pueden realizarse, a través del diálogo de búsqueda avanzada, consultas y búsquedas complejas aplicadas a los campos de QlikView, respetando la sintaxis de expresiones de script que permite QlikView. Este diálogo puede invocarse mediante la abreviatura de teclado CRTL+MAYÚS+F.

La expresión debe comenzar con un signo =, seguido de la expresión deseada. La expresión se evaluará para cada valor de campo de búsqueda mostrando los valores para los que la expresión devuelve un valor distinto de 0.

Un ejemplo de expresión de búsqueda es el que sigue:

`= sum(`*`patologia`*`)>sum(`*`experto`*`)`

Que realizada sobre un cuadro de lista que contenga el campo ‘estudio’, devolverá los estudios de arquitectura con un número de patologías asociadas mayor que el número de expertos asociados.

3.3.4 Objetos de hoja

QlikView maneja catorce tipos de objetos de hoja: cuadros de lista, cuadros de estadísticas, cuadros de selección múltiple, cuadros de tabla, gráficos, cuadros de entrada, cuadros de selecciones actuales, botones, objetos de texto, objetos de línea/flecha, objetos calendario/deslizador, objetos personalizados y objetos de marcador.

Pueden crearse objetos de hoja de tres tipos:

- Objetos de hoja de documento. Que se almacenan en el documento .qvw, estando siempre disponibles independientemente de si el archivo es abierto en local o desde un QlikView Server.

- Objetos de servidor personales. Disponibles sólo para usuarios autenticados y trabajando sobre un QlikView Server.

- Objetos de servidor compartidos. Disponibles sólo para usuarios autenticados y usuarios con permiso para visualizar el objeto, trabajando sobre un QlikView Server.

En el presente punto se detallarán las características de los objetos incluidos en la aplicación QlikView generada para el proyecto. Todos los objetos de la aplicación se han creado como objetos de hoja de documento, al no estar trabajando sobre un QlikView Server, ni tener motivos de seguridad para aplicar permisos a diferentes usuarios.

Cada objeto de hoja tiene propiedades inherentes a él, éstas pueden ser configuradas a través del diálogo *Propiedades* del menú *Objeto*.

3.3.4.1 Cuadro de lista

El cuadro de lista es el objeto de hoja más básico de QlikView. Se trata de una lista de todos los valores posibles de un campo. Es sobre el cuadro de lista, sobre el que se realizan las selecciones y las búsquedas explicadas en el punto 3.2.3. Sobre esas selecciones se pueden observar las relaciones e implicaciones entre campos, tal como se explico en el punto 3.1.2 Lógica Asociativa. La imagen 27 ilustra un cuadro de lista de la aplicación.

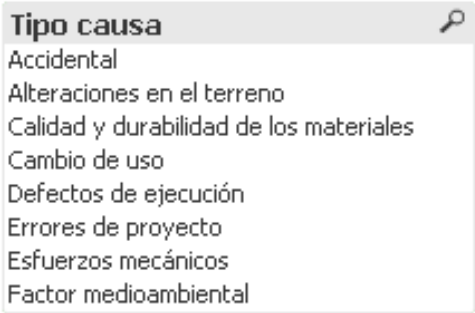


Ilustración 27. Cuadro de lista

Es posible ordenar los valores que muestra la lista, según diferentes criterios:

- Estado. Ordena los valores según su estado lógico, seleccionado, opcional o excluido. Éste es el criterio elegido en la aplicación.
- Expresión. Ordena los valores de campo según la expresión de ordenación deseada. Dicha expresión ha de cumplir con la sintaxis admitida por QlikView.
- Frecuencia. Ordena los valores según el número de veces que aparecen en el cuadro de lista.
- Texto. Ordena los valores por orden alfabético.
- Orden de carga. Ordena los valores por su orden de carga, si éstos han sido cargados en diferentes momentos de tiempo.

3.3.4.2 Cuadro de selección múltiple.

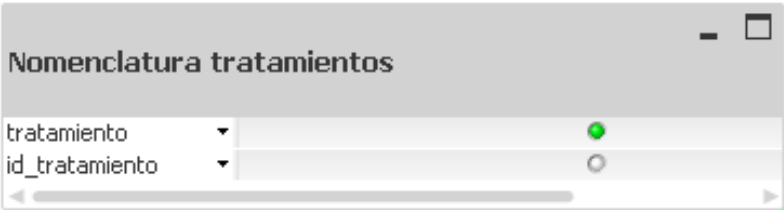


Ilustración 28. Cuadro de selección múltiple

El cuadro de selección múltiple permite mostrar diversos campos en un mismo objeto de hoja, como se muestra en la figura 28. Cada campo tiene un indicador con aspecto de bolita que da información del estado del campo. El indicador de color verde significa que hay valores seleccionados, blanco significa que hay valores opcionales y gris que no hay valores relacionados en la lista despegable.

Ilustración 29. Despliegue del cuadro selección múltiple

Haciendo click en uno de los campos del cuadro, se despliega la lista de valores del mismo, como se observa en la figura 29. En esta lista pueden realizarse selecciones únicas o múltiples, que siguen la misma codificación de colores. Se trata de selecciones como las que pueden hacerse en los cuadros de lista, por lo tanto selecciones de la lista desplegable pueden alterar la selección actual de valores en otros objetos de hoja.

En el diálogo de propiedades del cuadro, pueden configurarse aspectos como por ejemplo *Ignorar nulos,* para que no se muestren en la lista desplegable valores excluidos de la selección actual.

3.3.4.3 Cuadro de selecciones actuales.

Cuando hay muchas selecciones hechas sobre el documento, es difícil obtener una visión general de los datos. Para ello, el cuadro de selecciones actuales, como el que se muestra en la figura 30, contiene una lista de todas las selecciones realizadas sobre el documento, manteniendo para estas selecciones tres campos de información:

- Campos. Muestra los nombres de los campos sobre los que se hicieron selecciones, o búsquedas que llevaron a selecciones.

- Estado. Muestra el estado lógico del campo. Un campo puede des-seleccionarse desde este objeto haciendo click sobre su icono .

- Valores. Muestra los valores correspondientes a cada uno de los campos seleccionados.

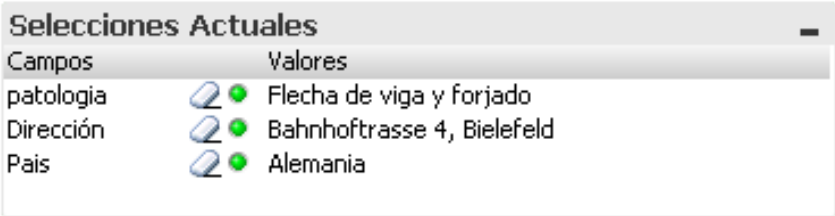


Ilustración 30. Cuadro de selecciones actuales

El cuadro de selecciones actuales, al igual que los demás objetos de hoja, se actualiza de manera dinámica a la vez que se realizan distintas selecciones, permitiendo una visión general permanente de las selecciones.

3.3.4.4 Botones.

Los botones en QlikView sirven para que se realice una acción al pulsar sobre ellos.

Para que el botón conlleve una acción asociada, ha sido necesario fijarla en la página acciones en el diálogo propiedades de botón.

Existen treinta y cinco comandos o acciones que pueden ser ejecutadas por un botón, que se clasifican en cinco grupos de acciones:

- Acciones de selección. Que ejecutan comandos de selección de campos por su estado lógico, bloquea, desbloquea o borra selecciones.

- Acciones de diseño. Ejecutan comandos como la activación y la restauración de objetos.

- Acciones de marcador. Ejecutan comandos que crean o reemplazan marcadores (objetos de hoja que sirven para mostrar y seleccionar marcadores de selección)

- Acciones de impresión. Ejecutan comandos que permiten imprimir objetos, hojas o informes.

- Acciones externas. Ejecutan comandos diversos como exportación, abrir una url, cerrar el documento, etc.

Los botones incluidos en la aplicación, como el que muestra la figura 31, tienen como objetivo eliminar las selecciones que estén realizadas sobre el documento, para aportar claridad o comenzar con nuevas búsquedas sobre datos.

Ilustración 31. Botón

3.3.4.5 Objeto de texto

El objeto de texto, mostrado en la figura 32, es utilizado en QlikView para mostrar información de texto o imágenes.

En la aplicación se ha utilizado este tipo de objeto para cargar la descripción de los tratamientos utilizados para las patologías que forman parte de la selección actual. Este objeto ha sido muy útil para tal fin, porque de no haberlo diseñado de esa forma, se habrían tenido que cargar los tratamientos en un objeto de cuadro de lista, y por la extensión del campo tratamiento, hubiera sido un objeto de difícil comprensión, ya que se mostrarían todos los tratamientos almacenados en la base de datos, y de esta manera se consigue que se muestren sólo los vinculados a la búsqueda actual del usuario.

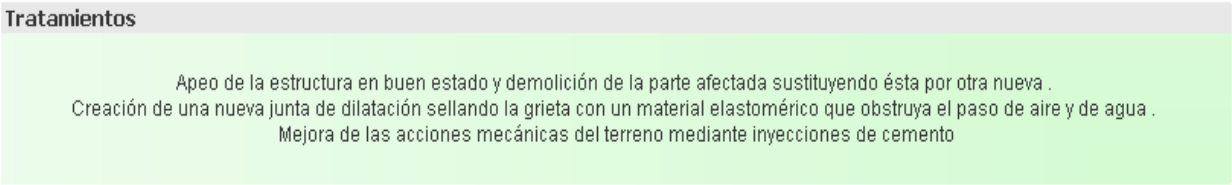


Ilustración 32. Objeto texto

3.3.4.6 Objeto calendario/deslizador

El objeto calendario/deslizador es otra manera de representar los valores de un campo determinada, representa una alternativa al cuadro de lista. Representa los valores de una manera más visual, siendo apropiado su uso sobre todo para campos de fecha, uso que se le ha dado en la aplicación realizada.

El objeto calendario/deslizador se llama así porque puede representarse como:

- Objeto calendario. Que muestra el campo como una vista de calendario clásico, estilo calendario de Windows.

- Objeto deslizador. Se trata del tipo elegido para la aplicación, muestra los valores de campo en un segmento de valores. Debe definirse una escala para este tipo de objeto, que defina la frecuencia de marcas de escala de los valores. Esto sirve para indicar el número de valores se quiere que aparezcan en el deslizador, cuántas escalas principales y cuántas escalas secundarias para cada escala principal. La figura 33 muestra el objeto deslizador utilizado en la aplicación, en el cual pueden seleccionarse diferentes fechas de realización de proyectos, que conllevaría la selección de los mismos en los cuadros de lista destinados al efecto.

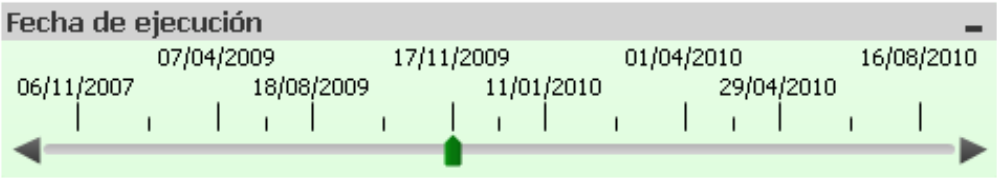


Ilustración 33. Objeto deslizador

3.3.4.7 Gráficos

Los gráficos son objetos de hoja que permiten mostrar resultados de funciones de manera compacta y visual. Estos resultados se calculan utilizando varios campos de las tablas de la base de datos, y sólo pueden ser mostrados a través de un gráfico.

Las selecciones que se realicen sobre los campos en los cuadros de lista activarán el cálculo correspondiente a la expresión del gráfico, modificando el aspecto del mismo. Pueden realizarse también si se desea, selecciones sobre los datos del propio gráfico, serán estas selecciones entonces las que modifiquen la selección actual de los cuadros de lista.

Existen dos categorías de gráficos en QlikView, gráficos de imagen y gráficos de tabla:

1. Gráficos de imagen.

Se trata de los gráficos propiamente dichos. QlikView permite la construcción de diez tipos diferentes de gráficos:

- Gráfico de barras. Se trata del tipo de gráfico más básico. Los valores de los campos se muestran como barras de diferentes alturas.

- Gráfico de dispersión. Presenta en cada uno de sus ejes los resultados de dos expresiones diferentes asociadas al gráfico. Resulta de utilidad para mostrar instancias asociadas a dos valores.

- Gráfico de líneas. Los valores se muestran como líneas entre puntos de valores, solamente como valores, o como combinación de ambos, líneas y valores. Este tipo de gráfico es muy útil para mostrar tendencias, evoluciones en los datos.

- Gráfico combinado. Permite mostrar los valores de una primera expresión como barras y los de una segunda como líneas.

- Gráfico de radar. Se llama así por su aspecto similar a una pantalla de radar. Se trata una versión del gráfico de líneas, en la que el eje-x se enrolla 360º.

- Gráfico de rejilla. Similar en aspecto al gráfico de dispersión, este tipo de gráfico traza los valores de dimensión en los ejes.

- Gráfico de indicador. Muestra el resultado de una única expresión, sin emplear dimensiones.

- Gráfico de bloques. Muestra los valores devueltos por la expresión como bloques de diferentes áreas

- Gráfico de tarta. Muestran la relación entre una dimensión y la expresión del gráfico.

- Gráfico de embudo. Similar en aspecto al gráfico de tarta, este gráfico representa los valores como diferentes segmentos con una anchura y altura proporcional al valor devuelto por la expresión del gráfico.

En la imagen 34 se ilustran los componentes básicos que aparecen en los gráficos de imagen realizados en la aplicación.

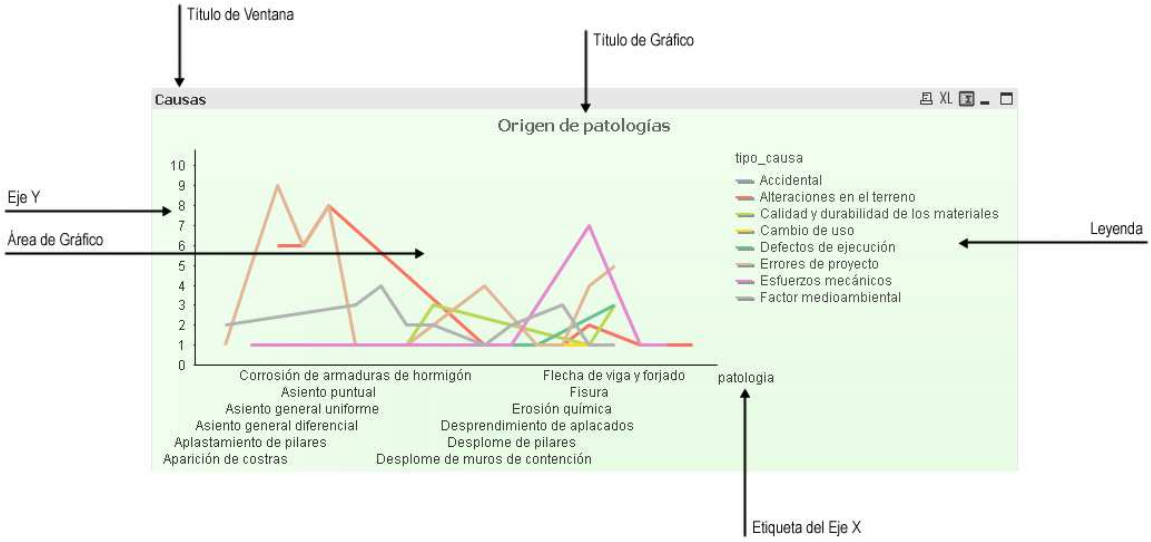


Ilustración 34. Gráficos de imagen

2. Gráficos de tabla.

Los gráficos de tabla sirven para mostrar los datos con un diseño de tabla, en vez de utilizar un diseño gráfico. Aunque guarden cierto parecido, los gráficos de tabla no son el mismo tipo de objeto de hoja que los cuadros de tabla. Los gráficos de tabla muestran valores asociados al cálculo de una expresión sobre una serie de campos, mientras que los cuadros de tabla muestran los valores que toman dichos campos en la base de datos, sin realizar ningún cálculo sobre los mismos.

Existen dos tipos de gráfico de tabla: Las tablas simples pivotantes y las tablas pivotantes. La diferencia entre ambas es que las tablas pivotantes muestran valores subtotales, mientras que las simples pivotantes no lo hacen.

En la imagen 35 se ilustran los componentes básicos que aparecen en los gráficos de tabla.

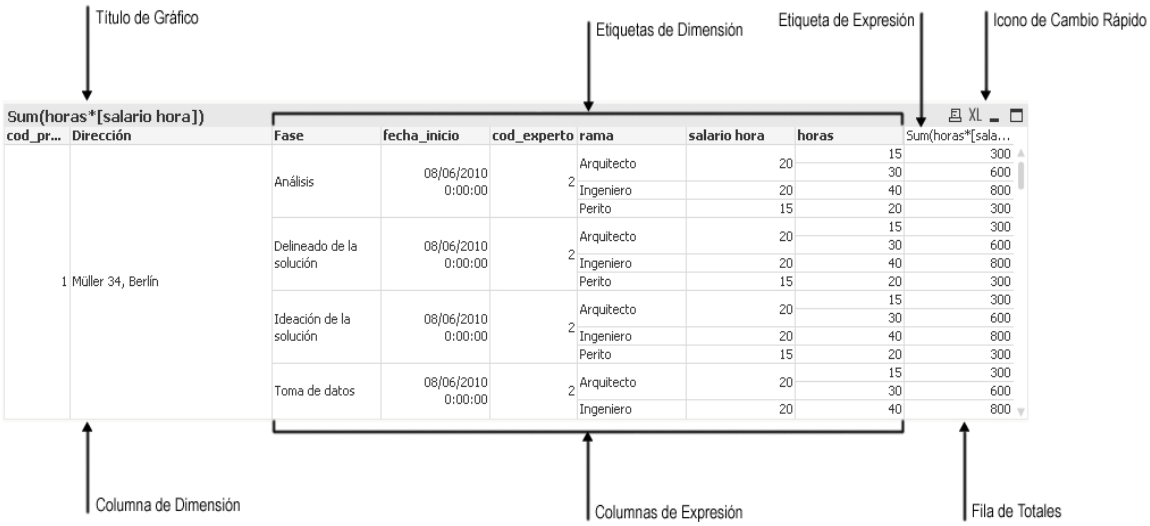


Ilustración 35. Gráficos de tabla

Cambio de estilo de gráfico

El icono de cambio de estilo de gráfico que aparece a un lado en el título del gráfico, indica que en un mismo objeto de hoja gráfico se han incluido varios tipos de gráficos, que muestran sus valores bajo la misma expresión de cálculo. Esto es útil para observar una misma información de diversas maneras, que puedan llevar al final a una mejor comprensión de la información presentada en el documento QlikView.

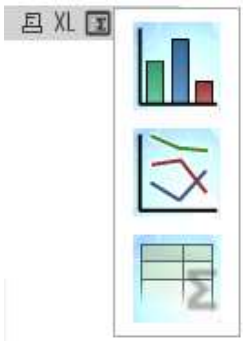


Ilustración 36. Cambio estilo de gráfico

Haciendo click con el botón derecho sobre el icono se desplegará la lista de los tipos de gráficos disponibles para el objeto de hoja. El gráfico cambiará de tipo al hacer click sobre el icono que lo representa. El desplegable de la figura 36 indica que el gráfico

activo puede visualizarse como un gráfico de barras, de líneas o como una tabla simple pivotante.

Expresiones de gráfico

Es necesaria la definición de una expresión de cálculo en los gráficos, para determinar los datos mostrados en el mismo. Esta expresión debe ser definida según la sintaxis de script admitida por QlikView. La expresión devolverá un número o cadena distinto de null.

Para cada gráfico se describirá su expresión asociada en el siguiente apartado:

Gráficos en la aplicación

Se detallan a continuación los gráficos empleados en la aplicación y la información que permite visualizar cada uno.

Gráfico 1. Tratamientos utilizados

Las dimensiones del gráfico1 son patología y tratamiento, y se encuentra ordenado por la primera de las dimensiones. Muestra por lo tanto cuántos tratamientos existen para cada una de las patologías registradas y cuáles son los identificadores de estos tratamientos.

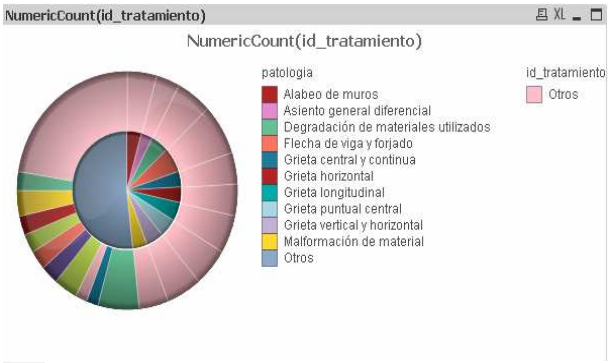


Ilustración 37. Gráfico 1

La expresión de cálculo de este gráfico es:

```
NumericCount (distinct id_tratamiento)
```

La función NumericCount devuelve el contador númerico del campo id_tratamiento.

El cualificador distinct delante del nombre del campo indica que para el cálculo no se tendrán en cuenta valores duplicados del tratamiento asociado a una patología, si éstos existieran.

Gráfico 2. Orígenes de patologías.

Las dimensiones del gráfico2 son patología y tipo de causa, y se encuentra ordenado por la primera de las dimensiones. Muestra información de los tipos de causa que originan una determinada patología.

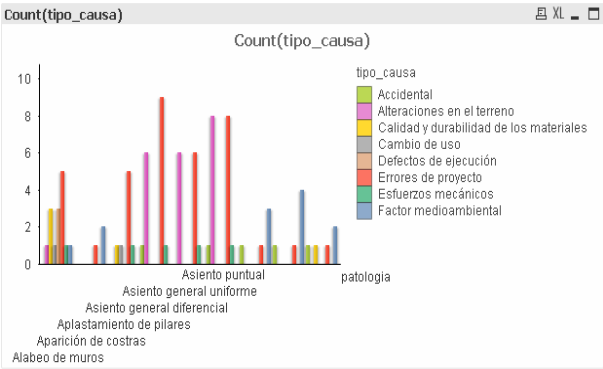


Ilustración 38. Gráfico 2

Permitiendo descubrir por ejemplo que para la patología 'Asiento general diferencial' los errores en el diseño del proyecto de la estructura son el factor principal desencadenante de la misma, seguido de las alteraciones en el terreno.

La expresión de cálculo de este gráfico es:

```
Count (tipo_causa)
```

La función Count devuelve el contador de totales agregado de los valores del campo tipo_causa sobre las dimensiones del gráfico.

Gráfico 3. Patologías por países.

Las dimensiones del gráfico3 son patología y país, y se encuentra ordenado por la primera de las dimensiones. Muestra información de las patologías que se originan en los diferentes países en los que la organización tiene estudios.

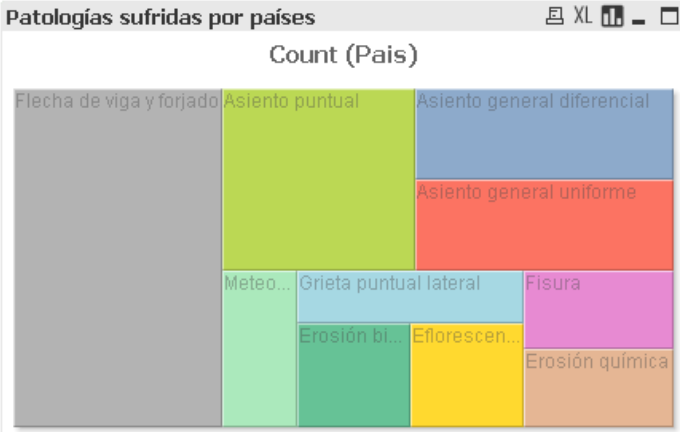


Ilustración 39. Gráfico 3

Permitiendo descubrir por ejemplo que en Alemania, la patología más veces tratada es 'Flecha de viga y forjado'. Este conocimiento podría ser relacionado entonces con las causas, para determinar que uno de los tipos de causa principales que la originan son los esfuerzos mecánicos. De esta manera se va deduciendo el conocimiento.

La expresión de cálculo de este gráfico es:

```
Count (Pais)
```

Tiene opción de cambio de estilo rápido, por lo que información mostrada en él puede verse como gráfico de bloques, mostrado en la imagen 39, o bien como gráfico de barras, o de embudo.

Gráfico 4. Proyectos por patología

La dimensione del gráfico4 es id_patología. Este gráfico muestra los proyectos que se llevan a cabo por cada patología.

Permitiendo descubrir, de los proyectos realizados, cuál es la patología más recurrente.

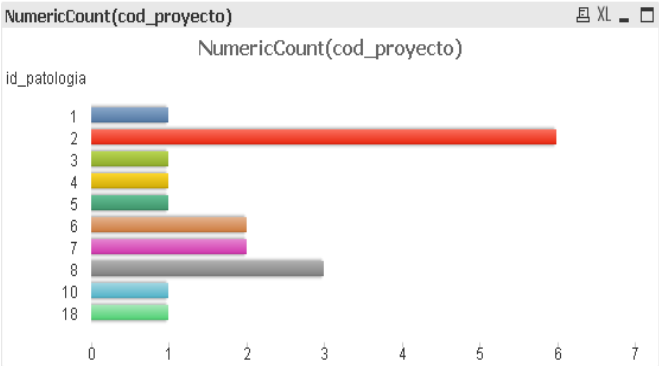


Ilustración 40. Gráfico 4

La expresión de cálculo de este gráfico es:

```
NumericCount(cod_proyecto)
```

La función NumericCount devuelve el contador númerico del campo cod_proyecto.

Gráfico 5. Presupuesto

El gráfico 5 no posee dimensión de entrada. Muestra información del coste total de realización de los proyectos que conformen la selección actual.

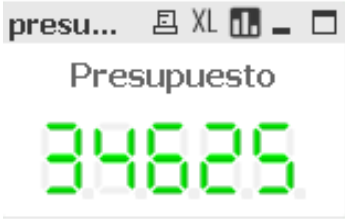


Ilustración 41. Gráfico 5

La expresión de cálculo de este gráfico es:

```
Sum (horas*[salario hora])
```

La función Sum devuelve la suma de la expresión entre paréntesis, que

consiste en el salario por hora de los recursos humanos implicados en el proyecto multiplicado por las horas de trabajo que realiza cada uno de ellos.

Gráfico 6. Asignación de proyectos

Las dimensiones del gráfico 6 estudio y proyecto se encuentra ordenado por la segunda de las dimensiones. Muestra información de proyectos que ha llevado a cabo cada estudio de la organización.

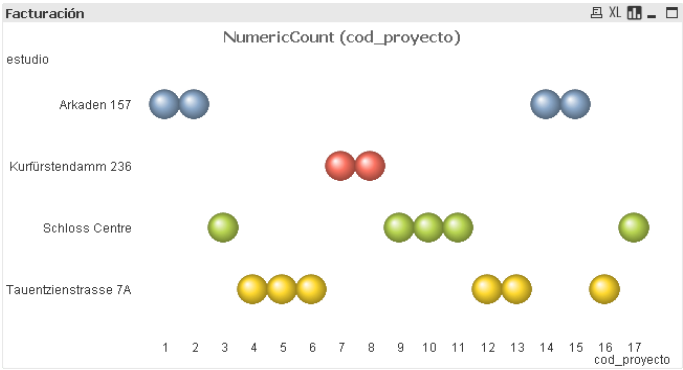


Ilustración 42. Grafico 6

Combinado con la información que desprende el gráfico anterior, pueden descubrirse cuáles son los estudios que mayor número de proyectos realizan y que más rentables son.

La expresión de cálculo de este gráfico es:

```
NumericCount (cod_proyecto)
```

Tiene opción de cambio de estilo rápido, por lo que información mostrada en él puede verse como gráfico de dispersión, mostrado en la imagen 41, bien como gráfico de barras.

Gráfico 7. Relación entre patologías

Las dimensiones del gráfico 7 son id_patología y patologías asociadas, Muestra información sobre qué patologías se encuentran asociadas entre sí.

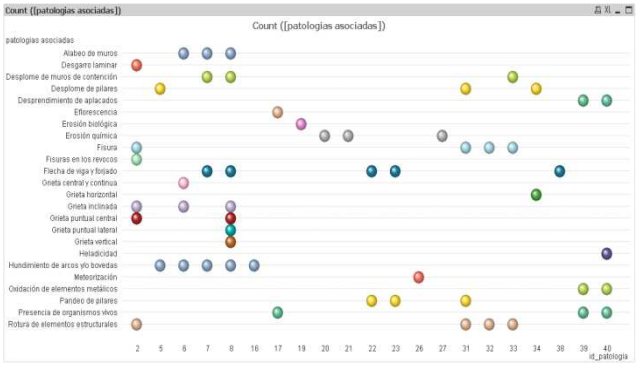


Ilustración 43. Gráfico 7

La expresión de cálculo de este gráfico es:

```
Count ([patologías asociadas])
```

4. Planificación y recursos

Este capítulo presenta, en un diagrama Gantt, la secuencia de tareas realizadas para la consecución del presente proyecto, junto con la lista de recursos necesarios para la realización del mismo.

4.1 Recursos

A lo largo del desarrollo del proyecto se han empleado los siguientes recursos:

- 1 ordenador portátil con S.O Windows XP Profesional 32 bits.
- Disco duro IDE USB de 300 GB, para el almacenamiento de backup´s.
- Conexión internet móvil con una velocidad 1Gbps. Se empleará software de protección antivirus Kaspersky Internet Security.
- Microsoft Office 2007. Suite de aplicaciones entre las que destacan la herramienta de procesamiento de texto Microsoft Word y la herramienta de desarrollo de bases de datos Access.
- Microsoft Project Proffesional 2010.

- Photoshop. Aplicacion para el diseño de imágenes incluidas en el presente documento.
- Adobe Reader.
- QlikView 9.0 Personal Edition.

4.2 Planificación

Las tareas realizadas pueden dividirse en dos grandes grupos, como puede también observarse en el diagrama Gantt de la imagen 44:

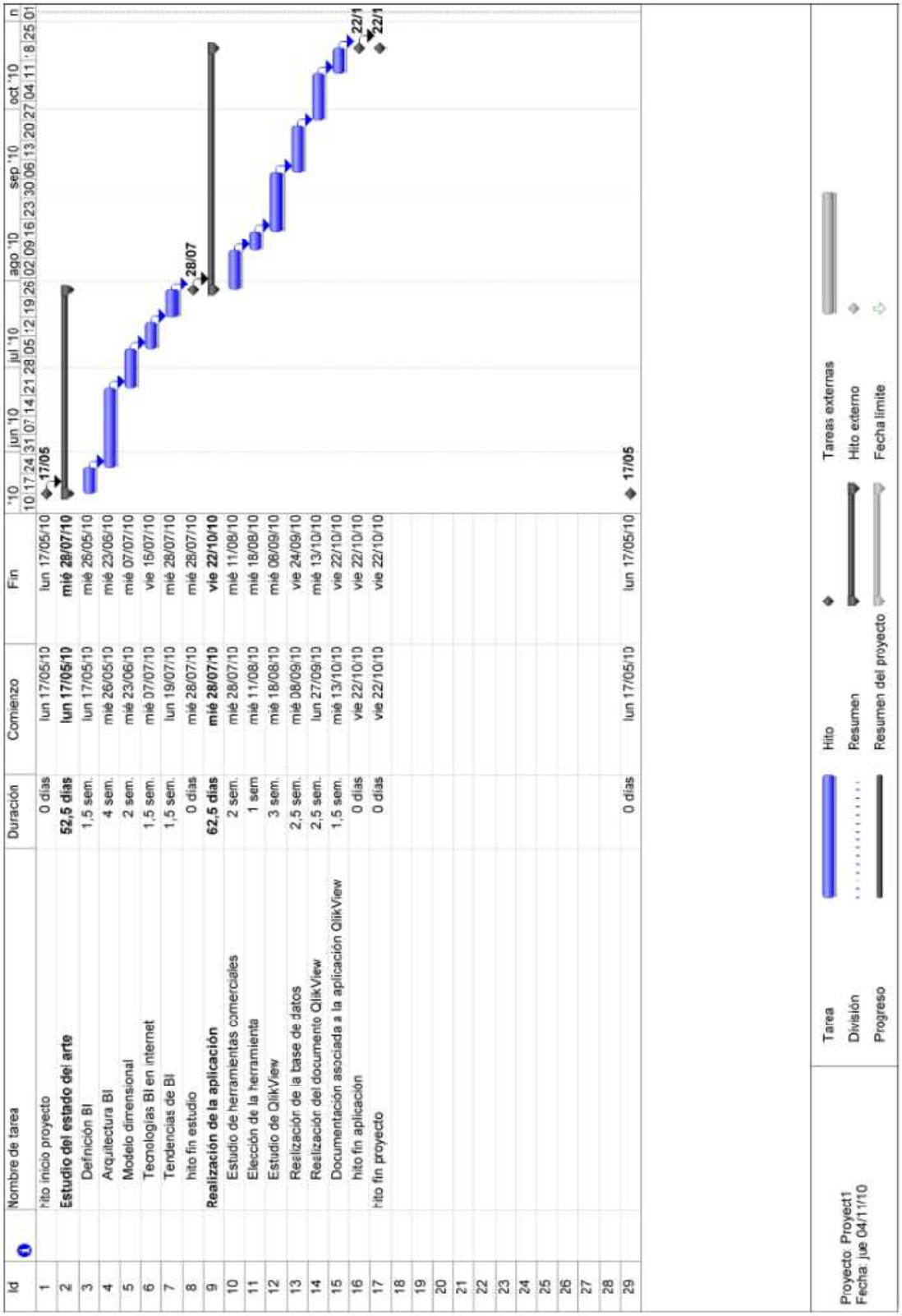
Estudio de Business Intelligence:

- Estudio del concepto de BI
- Estudio de la arquitectura BI
- Estudio del modelo dimensional
- Uso de tecnologías BI en Internet
- Estudio de tendencias del BI

Realización de la Aplicación:

- Estudio de herramientas comerciales de BI
- Elección de la herramienta
- Estudio del manejo de QlikView y características en su arquitectura
- Realización de la base de datos BDPatologias.access
- Realización del documento BI QlikView.
- Documentación asociada al desarrollo del documento QlikView.

Diagrama de Gantt



Id	Nombre de tarea	Duración	Comienzo	Fin
1	hito inicio proyecto	0 días	lun 17/05/10	lun 17/05/10
2	**Estudio del estado del arte**	**52,5 días**	**lun 17/05/10**	**mié 28/07/10**
3	Definición BI	1,5 sem.	lun 17/05/10	mié 26/05/10
4	Arquitectura BI	4 sem.	mié 26/05/10	mié 23/06/10
5	Modelo dimensional	2 sem.	mié 23/06/10	mié 07/07/10
6	Tecnologías BI en internet	1,5 sem.	mié 07/07/10	vie 16/07/10
7	Tendencias de BI	1,5 sem.	lun 19/07/10	mié 28/07/10
8	hito fin estudio	0 días	mié 28/07/10	mié 28/07/10
9	**Realización de la aplicación**	**62,5 días**	**mié 28/07/10**	**vie 22/10/10**
10	Estudio de herramientas comerciales	2 sem.	mié 28/07/10	mié 11/08/10
11	Elección de la herramienta	1 sem	mié 11/08/10	mié 18/08/10
12	Estudio de QlikView	3 sem.	mié 18/08/10	mié 08/09/10
13	Realización de la base de datos	2,5 sem.	mié 08/09/10	vie 24/09/10
14	Realización del documento QlikView	2,5 sem.	lun 27/09/10	mié 13/10/10
15	Documentación asociada a la aplicación QlikView	1,5 sem.	mié 13/10/10	vie 22/10/10
16	hito fin aplicación	0 días	vie 22/10/10	vie 22/10/10
17	hito fin proyecto	0 días	vie 22/10/10	vie 22/10/10
29		0 días	lun 17/05/10	lun 17/05/10

Ilustración 44. Diagrama de Gantt

4.3 Cálculo de costes

Conforme a lo expuesto, el presupuesto que hubiese conllevado el presente proyecto sería el que se presenta en la tabla 13.

Coste de los recursos en €	
Ordenador portátil	700
Disco duro IDE USB de 300 GB	60
Conexión internet móvil 1Gbps.	160
Licencia Kaspersky Internet Security.	49.95
Microsoft Office 2007.	155
Microsoft Project Proffesional 2010.	762.07
Photoshop.	497.3
Adobe Reader.	0
QlikView 9.0 Personal Edition.	0
SUB TOTAL	**2383.87**
Coste recursos humanos en €	
Estudio de BI	52´5 días * 8 hrs/día * 25€/hora = 10500€
Realización de la aplicación	62´5 días * 8 hrs/día * 25€/hora = 12500€
SUB TOTAL	**23000**
TOTAL	**25383.87**

Tabla 13. Cálculo de costes

5. Conclusiones y Futuros Trabajos

5.1 Conclusiones

Los sistemas Business Intelligence facilitan notablemente el acceso a la información presentándola directamente para la toma de decisiones posterior por parte de sus usuarios.

Esto lleva a las empresas a conocer y manejar mejor la información de la que disponen, facilitando así la compresión global de la empresa, sus funciones, su productividad, sus objetivos, su estado y su estrategia. Hecho muy importante, ya que cuando el personal no conoce toda la información al respecto, puede equivocarse en las decisiones de negocio, y como dijo Peter Drucker[9], *"no hay nada peor que cometer bien los errores"*.

El BI lleva a una conclusión interesante (relacionando con otras asignaturas vistas en la carrera) y es que de su utilización podría darse un aplanado en los organigramas de las empresas.

[9] Peter Ferdinand Drucker. Escritor y gurú de la gestión, autor de múltiples obras reconocidas mundialmente sobre temas referentes a la gestión de las organizaciones, sistemas de información y sociedad del conocimiento.

Típicamente las empresas han tenido una organización bastante jerarquizada en forma de pirámide, debido en gran parte a que debe haber una persona encargada de la toma de decisiones, para que el desarrollo de los procesos que en la empresa se dan sean uniformes y se garantice así la repetitividad y la calidad. Las decisiones las toma por lo tanto la persona que posee la información, que es la persona que se encuentra más arriba en el organigrama (llegándole la información de las capas inferiores).

La nueva propuesta de aplicar plataformas BI dota a los empleados de la organización de la información necesaria para poder decidir. Esto llevaría, como hemos dicho a un aplanado de la estructura de la organización, disminución de los costes, disminución del tiempo en la realización de los procesos de la empresa al eliminar tiempos de espera. Concluyendo, el BI ayudaría en gran medida a los proyectos de reingeniería que puedan plantearse en una organización.

5.2 Futuros trabajos.

Podrían realizarse como futuros trabajos que continuarán la línea de éste:

- Desarrollo de una aplicación BI completa, y montada sobre un servidor, tomando como origen de datos la base de datos alojada en la universidad sobre patrones de producto.

- Seguimientos de las tendencias previstas para el Business Intelligence, explicadas en el punto 2.8 del presente documento, y realización de un nuevo estudio si fuese necesario.

Anexo: Jerarquía DIKW.

DIKW (**d**ata, **i**nformation, **k**nowledge, **w**isdom), también conocida como la jerarquía de la sabiduría, es modelo jerárquico que se representa a menudo en forma de pirámide con los *datos* en su base y la *sabiduría* en su ápice y en la que cada capa agrega ciertas cualidades superior la anterior. Fue propuesta por Milan Zeleny (1987) y Russel Ackoff (1989) y propone una forma de entender la relación entre los conceptos de *datos*, *información*, *conocimiento* y *sabiduría*.

Estos términos pueden usarse, si entablamos una conversación informal, de manera indistinta, pero no se refieren a lo mismo. Me ha parecido importante incluir un anexo explicando la diferencia, ya que son términos que se han utilizado bastante en el presente documento.

Datos

El primer nivel de este modelo es formado por datos. Un dato es la unidad mínima semántica y se corresponde con elementos primarios de información. También se puede ver como un conjunto discreto de factores objetivos sobre un hecho real. Los datos vienen bajo la forma de observaciones y medidas, al observar algo, al medirlo, estamos captando datos.

Pero los datos son elementos aislados, no dicen nada del por qué de las cosas, carecen de un significado útil, es decir, que simplemente existen y por sí mismos no tienen relevancia y no son capaces de orientarnos a ninguna decisión. Los datos pueden ser números, palabras, sonidos o imágenes que no son necesariamente organizados.

Lo cual no quiere decir que los datos no sean importantes, todas las organizaciones necesitan y almacenan cantidad de datos (los bancos, las compañías de seguros y la seguridad social son unos clarísimos ejemplos).

Ejemplo:

‘está lloviendo’: Esto es un dato, representa un hecho.

Información

Se puede dar un significado a los datos al crear una relación entre ellos. Esto último se convierte en lo que se conoce como información, que forma el siguiente nivel del modelo DIKW.

Podemos definir información como un conjunto de datos procesados. La información es creada analizando relaciones y conexiones entre los datos. Es capaz de contestar a las preguntas de quién/qué/dónde/cuándo/por qué. A diferencia de los datos, la información tiene significado (relevancia, propósito y contexto).

La palabra 'informar' originalmente significa 'dar forma a' y es porque la información es capaz de formar a quién la recibe, impactando sobre sus juicios de valor y comportamientos. Reduce su incertidumbre sobre el dato y es capaz de meter a la persona en contexto.

La información es un mensaje, existe un emisor, receptor y está siempre organizada para algún propósito. Aunque estrictamente hablando es el receptor y no el emisor quién decide si el mensaje que ha recibido es realmente información, si realmente le informa, 'si cambia su forma'.

Así por ejemplo un fichero de texto lleno de registros consecutivos delimitados por un carácter puede tener sentido para el propietario del fichero, pero puede que sea juzgado como 'ruido' por otras personas.

Transformamos los datos en información cuando les aportamos significado. Para este fin existen varios métodos:

- Contextualizando: Cuando sabemos en qué contexto y para qué propósito se generan los datos.
- Categorizando: Conocemos la unidad de medida, la manera de clasificar los datos, entonces podremos agruparlos en diferentes clases.
- Calculando: También podemos obtener información de los datos al procesarlos matemática o estadísticamente.
- Corrigiendo: Podemos obtener información también al eliminar inconsistencias o errores que existieran en los datos.

- Condensando: Por lo general las empresas suelen almacenar muchos datos y puede ser complicado identificar cuáles son relevantes. Por el hecho de condensar, resumir de forma más concisa esos datos, podemos obtener información.

Información = Datos + Contexto (añadir valor) + Utilidad (reducir incertidumbre)

Ejemplo:

'La temperatura bajó 15 grados y luego empezó a llover', esto es información, la información encarna la comprensión de una relación de algún tipo, particularmente de causa-efecto.

Conocimiento

Una colección adecuada de información nos lleva al siguiente nivel en el modelo, que es el conocimiento.

El conocimiento es algo más rico y profundo que los datos y la información. Para Davenport y Prusak el conocimiento es "una mezcla de experiencia, valores, información y cierto 'know-how' que sirve de marco para la incorporación de nuevas experiencias e información, y es útil para la acción".

El conocimiento contesta a la pregunta "cómo". El conocimiento se origina en la mente de las personas conocedoras.

Se deriva de la información, así como la información se deriva de los datos. Pero para que la información se convierta en conocimiento, las personas tenemos que hacer procesos como:

- Comparación: Comparamos unos elementos con otros y de ahí sacamos conclusiones.

- Búsqueda de conexiones.

Siguiendo con el ejemplo anterior:

‘Si la humedad es muy alta y la temperatura disminuye considerablemente la atmósfera es a menudo poco probable que sea capaz de mantener la humedad por lo que llueve.’

El conocimiento representa un patrón que conecta y generalmente provee un alto nivel de previsibilidad en cuanto a lo que se describe o lo que sucederá después.

Sabiduría

La información debe ser asimilada y procesada consciente o inconscientemente por la persona que la recibe, para que se convierta en conocimiento. La sabiduría se crea entonces con el uso del conocimiento, con la comunicación de los usuarios del conocimiento, y a través reflexión. Es un proceso de extrapolación de las personas que requiere la función mental que llamamos juicio. La sabiduría contesta a las preguntas “por qué” y “cuándo”. Se ocupa del futuro, pues toma implicaciones y efectos retrasados en consideración.

La sabiduría es la capacidad de tomar la mejor decisión posible basada en el conocimiento.

Para concluir con el ejemplo anterior podríamos decir que lo siguiente es sabiduría:

‘llueve porque llueve. Y esto abarca la comprensión de todas las interacciones que ocurren entre la lluvia, evaporación, corrientes de aire, los gradientes de temperatura, los cambios, y con lluvia.

Se resumen los conceptos en la ilustración 45.

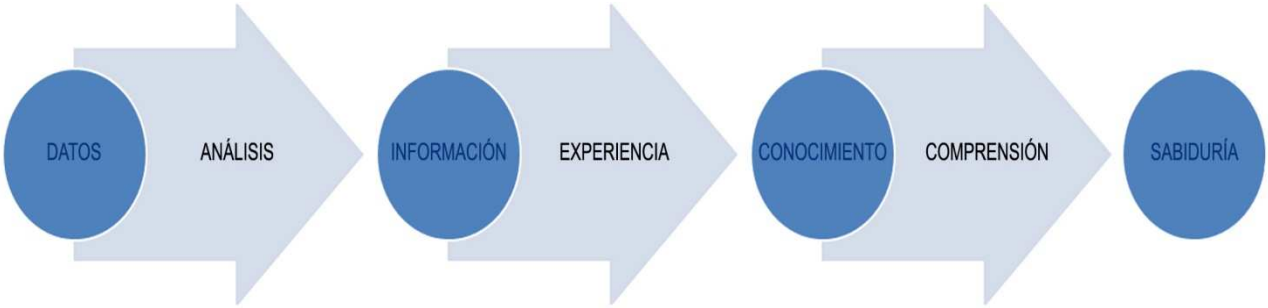


Ilustración 45. Jerarquia DIKW

Bibliografía

Noticias/Artículos

- [Rice,2004] Matthew B. Rice, "*Business Intelligence* tools can help turn out savings in core cost areas" , Managed Healthcare Alliance, Marzo 2004.
- [B. Inmon, 2006]. Visto en http://www.b-eye-network.com
- [M.Beyer et al, 2006] de Mark Beyer y Donald Feinberg "The Basic of datawarehousing", 2006.
- [Wayne W., 2002] Wayne W. Eckerson, "Data quality and the bottom line", TDWI, Report Series 2002.
- [J.Watson, 2006] Hugh. J. Watson, "Recent Development in datawarehousing" publicado en http://www.terry.uga.edu/~hwatson/dw_tutorial.ppt , 2006
- [QlikTeach , 2006] D. Vesset ,"QlikTech´s Approach to Business Intelligence: Keep it Simple and Flexible", Julio 2006.
- [InformationWeek, 2009] "Next-Gen BI is here" Information Week, InformationWeekanalytics.com, Septiembre 2009.
- [barc, 2007] BARC. Business Aplication Research Center. http://www.bi-verdict.com/
- [Wise, 2007] Lyndsay Wise ,"Cinco pasos para lograr un proyecto de Business Intelligence exitoso" ,Junio 2007.
- [Turban, 1995] Turban, E. (1995). Decision support and expert systems: management support systems. Englewood Cliffs, N.J., Prentice Hall.
- [W. Eckerson et al. 2005] "Business Intelligence: Competir con la información". Josep Lluís Cano.
- [CIO Insight] "The future of I.T: What´s on tap for 2007 and beyond", Enero 2007

- [Stephen B.] Stephen Brobst Chief Technology Officer - Teradata Corporation, Septiembre 2010.
- "A Business Intelligence System" *IBM Journal*. Julio 2008
- Data Quality for BI http://www.trilliumsoftware.com/home/solutions/enterprise/datawarehouse_bi.aspx
- "La gestión del conocimiento. Su importancia como recurso estratégico para la organización" Sanguino, R, 2003.
- "Repositorios OLAP". Carlos Hurtado Larrain. Presentación en el Congreso de Ciencia y Tecnología Computacional, Tuluá, Abril 2005.
- "OLAP and Datamining: Brigging the Gap" http://www.datamining.com/bridge.htm
- "Cómo diseñar grandes variables en bases de datos multidimensionales". Manuel de la Herrán Gascón. Abril 2001.
- "Datamining: Oráculos de sinlencio. Buceando en un mar de información". Sergio A. Moriello. Noviembre 1999.
- "The engineering of knowledge-based systems theory and practice", Douglas D. Dankel, Avelino J Gonzales. Prentice-Hall, 1993.
- "Enterprise BI: Strategies and technologies for deploying BI on an enterprise scale" Wayne Eckerson, Agosto 2005.
- "The Balanced Scorecard: Translating Strategy into Action", Harvard Business School Press, Boston, 1996.
- "The Datawarehouse lifecycle toolkit" Ralph Kimball, Laura Reeves. Wliley computer publishing, 1998.
- "Patología y técnicas de intervención en estructuras arquitectónicas" Monjo Carrió y Luis Maldonado Ramos "Ediciones munilla-lería, 2001.

Enlaces web

- www.gartner.com.
- http://www.bi-spain.com
- SAP BI: www.sap.com
- QlikView: http://www.qlikview.com
- Microsoft BI: http://www.microsoft.com/bi/
- Oracle BI: http://www.oracle.com/
- Targit: http://www.targit.com
- www.nase-it.com
- www.todobi.blogspot.com
- www.BIScorecard.com
- Decisión:

  http://www.monografias.com/trabajos12/decis/decis.shtml
- Conocimiento, datos, información y sabiduría:

  http://www.gestiondelconocimiento.com/conceptos_diferenciaentredato.htm//

  http://itilv3.osiatis.es/proceso_mejora_continua_servicios_TI/DIKW.php
- Índice de referencias sobre BI y gestión del conocimiento

  www.bidigital.com/ci.index.html